UNIVERSITATEA DIN BUCUREŞTI

FACULTATEA DE PSIHOLOGIE ŞI ŞTIINȚELE EDUCAȚIEI

DEPARTAMENTUL DE ÎNVĂȚĂMÂNT LA DISTANȚĂ

BAZELE TEORETICE ALE EVALUARII PSIHOLOGICE

ANUL II

Prof. univ. dr. NICOLAE MITROFAN

Universitatea din Bucureşti
Editura CREDIS
2008

Copyright © DEPARTAMENT ID 2008

Acest material este destinat uzului studenţilor Universităţii din Bucureşti, forma de învăţământ la distanţă.
Conţinutul cursului este proprietatea intelectuală a autorului/autorilor; designul, machetarea şi transpunerea în format electronic aparţin Departamentului de Învăţământ la Distanţă al Universităţii din Bucureşti.

Universitatea din Bucureşti
Editura CREDIS
Bd. Mihail Kogălniceanu, Nr. 36-46, Corp C, Etaj I, Sector 5
Tel: (021) 315 80 95; (021) 311 09 37, 031 405 79 40, 0723 27 33 47
Fax: (021) 315 80 96
Email: credis@credis.ro
Http://www.credis.ro

Copyright © DEPARTAMENT ID 2008

Module:

Modulul 1 – Situaţia psihodiagnosticului în lume şi în ţara noastră

Modulul 2 – Testul psihologic ca mijloc de evaluare

Modulul 3 - Caracteristicile psihometrice ale testului psihologic

Modulul 4 - Analiza itemului

Modulul 5 – Etalonarea şi standardizarea testelor

Modulul 1 – Situaţia psihodiagnosticului în lume şi în ţara noastră

Unitatea de învăţare 1: Precizări prealabile privind evaluarea psihologică

Unitatea de învăţare 2: Contribuţii la dezvoltarea psihodiagnosticului în secolul XX

Unitatea de învăţare 3: Tendinţe şi direcţii de dezvoltare a psihodiagnosticului la începutul mileniului III.

Unitatea de învăţare 4. Situaţia psihodiagnosticului din România în perioada actuală

Modulul II – Testul psihologic ca mijloc de evaluare

Unitatea de învăţare 1: Testul psihologic

Unitatea de învăţare 2: Clasificarea testelor psihologice

Unitatea de învăţare 3: Testul psihologic şi Codul deontologic

Modulul III – Caracteristici psihometrice ale testului psihologic

Unitatea de învăţare 1: Ce sunt caracteristicile psihometrice ?

Unitatea de învăţare 2: Fidelitatea testului

Unitatea de învăţare 3: Validitatea testului

Unitatea de învăţare 4: Testul psihologic şi selecţia profesională

Modulul IV – Analiza itemului

Copyright © DEPARTAMENT ID 2008

Unitatea de învățare 1: Ce este analiza itemului ?

Unitatea de învățare 2: Gradul de dificultate al itemului

Unitatea de învățare 3: Capacitatea de discriminare a itemului

Modulul V – Etalonarea şi standardizarea testelor

Unitatea de învățare 1: Definirea conceptelor şi etapele construirii testului

Unitatea de învățare 2: Modele de scale utilizate în etalonare şi standardizare

Copyright © DEPARTAMENT ID 2008

MODULUL I

SITUATIA PSIHODIAGNOSTICULUI ÎN LUME SI ÎN ȚARA NOASTRĂ

Copyright © DEPARTAMENT ID 2008

Bazele teoretice ale evaluării psihologice

Module:

Modulul 1 – Situaţia psihodiagnosticului în lume şi în ţara noastră

Modulul 2 – Testul psihologic ca mijloc de evaluare

Modulul 3 - Caracteristicile psihometrice ale testului psihologic

Modulul 4 - Analiza itemului

Modulul 5 – Etalonarea şi standardizarea testelor

Modulul 1 – Situaţia psihodiagnosticului în lume şi în ţara noastră

Unitatea de învăţare 1: Precizări prealabile privind evaluarea psihologică

Unitatea de învăţare 2: Contribuţii la dezvoltarea psihodiagnosticului în secolul XX

Unitatea de învăţare 3: Tendinţe şi direcţii de dezvoltare a psihodiagnosticului la începutul mileniului III.

Unitatea de învăţare 4. Situaţia psihodiagnosticului din România în perioada actuală

Modulul II – Testul psihologic ca mijloc de evaluare

Unitatea de învăţare 1: Testul psihologic

Unitatea de învăţare 2: Clasificarea testelor psihologice

Unitatea de învăţare 3: Testul psihologic şi Codul deontologic

Modulul III – Caracteristici psihometrice ale testului psihologic

Unitatea de învăţare 1: Ce sunt caracteristicile psihometrice ?

Unitatea de învăţare 2: Fidelitatea testului

Unitatea de învăţare 3: Validitatea testului

Copyright © DEPARTAMENT ID 2008

Unitatea de învățare 4: Testul psihologic şi selecția profesională

Modulul IV – Analiza itemului

Unitatea de învățare 1: Teoria răspunsului la item

Unitatea de învățare 2: Gradul de dificultate al itemului

Unitatea de învățare 3: Capacitatea de discriminare a itemului

Modulul V – Etalonarea şi standardizarea testelor

Unitatea de învățare 1: Definirea conceptelor

Unitatea de învățare 2: Etapele etalonării şi standardizării testului psihologic

Unitatea de învățare 3: Modele de scale utilizate în etalonare şi standardizare

Copyright © DEPARTAMENT ID 2008

Câteva precizări prealabile....

Iată, aşadar, o nouă disciplină cu care veţi face cunoştinţă încă de la începutul anului II. Ea are un titlu tentant, desigur....evaluarea psihologică a individului uman....deci, vom putea să măsurăm pe cineva şi să aflăm cum „stă" în legătură cu diferite componente ale personalităţii. Vom şti cât este de inteligent, cât este de creativ, cât este de emotiv, cât este de sociabil, cât este de echilibrat, ce fel de temperament are, ce trăsături caracteriale are, cât de dezvoltate îi sunt anumite aptitudini etc.,etc. Putem să ne pronunţăm asupra cuiva în legătură cu coeficientul de inteligenţă (QI), coeficientul de dezvoltare (QD), coeficientul de creativitate (QKre), coeficientul emoţional (QE), coeficientul de personalitate (QP) etc. Păi, nu este formidabil să avem asemenea posibilităţi noi psihologii ? Evident că da şi pentru a face diverse evaluări putem apela la mai multe metode şi mijloace, însă noi ne vom concentra , mai ales, asupra testelor psihologice. Nu există sector al activităţii umane în care să nu fie necesară realizarea evaluării psihologice cu ajutorul testelor. Acestea sunt extrem de multe (unii autori apreciază că ar fi vorba de zeci de mii în lume), însă ele pot fi grupate în diferite categorii, în funcţie, în special de obiectivul urmărit în evaluare. Există însă şi anumite pericole...dar nu dorim să speriem pe nimeni încă de la început. Un pericol ar fi cel legat de utilizarea necorespunzătoare a testului. Ar fi ca şi cum am folosi un cântar defect pentru a stabili greutatea cuiva. Fiind defect, într-o zi ar arăta o anumită greutate, în altă zi, o greutate total diferită. Revenind în domeniul psihologiei, folosind, de exemplu, în mod greşit un test de inteligenţă, s-ar putea ca, la o primă testare, subiectul să apară ca fiind genial şi, la o altă testare, el să apară ca fiind deficient intelectual. Un alt pericol ar fi cel al, zicem noi, *delegării de răspundere*, adică psihologul conferă putere absolută testului, el neavând niciun fel de implicare în analizarea şi, mai ales, interpretarea rezultatelor. De aceea, un mare psiholog, pe nume L. Szondi, a prevenit pe toţi psihologii, din toate domeniile aplicative, spunându-le: ***„face mai mult un psiholog fără teste, decât mai multe teste fără psiholog".*** Poate că nu înţelegeţi exact ce a vrut să spună acest autor celebru, dar eu vă rog mult să-i reţineţi spusele, deoarece, dacă veţi practica psihologia, va trebui, cu siguranţă, să vi le amintiţi mereu. Revenind la pericole, exită şi multe altele, dar nu dorim să le expunem pe toate încă de pe acum. Ele vor reieşi după ce veţi parcurge toate modulele şi, mai ales, după ce vă veţi întâlni direct cu diferite categorii de teste. Am mai avea, însă o mare rugăminte. Să nu daţi uitării ce aţi învăţat în anul I, mai ales la *statistica aplicată în psihologie*. Ştim, ştim foarte bine că cei care vin spre domeniul psihologiei nu s-au manifestat prea pozitiv faţă de matematică şi, în general, faţă de ştiinţele exacte. Din nefericire însă pentru ei, la psihologie, încă din anul I apare....statistica iar despărţirea de ea este considerată de unii ca fiind pentru totdeauna. Evaluarea psihologică, psihodiagnosticul, mai ales, presupune cunoaşterea şi utilizarea unor elemente de statistică, aşa că, rugăm încă o dată să căutaţi sursele de informare şi să vă reactualizaţi cunoştinţele privind statistica aplicată în psihologie. Şi dacă ne-am înţeles până aici, vă propunem să ne apropiem de modulele pe care le oferim, începând, desigur, cu MODULUL I. Vă dorim mult succes !!!

Copyright © DEPARTAMENT ID 2008

Modulul 1 – Situația psihodiagnosticului în lume şi în țara noastră

Unitatea de învățare nr. 1

Precizări prealabile privind evaluarea psihologică

Cuprins:

1.1. Delimitări conceptuale

1.2. De când a început, de fapt, evaluarea psihologică ?

1.3. Cine este considerat părintele psihodiagnosticului ?

Obiective:

La sfârşitul acestei unități de învățare studenții vor fi capabili să:

- definească corect noțiunea de evaluare psihologică
- să diferențieze evaluarea psihologică ştiințifică de cea empirică
- cunoască ce criterii utilizăm pentru a stabili marile perioade de evoluție a evaluării psihologice
- cunoască unele dintre cele mai vechi preocupări privind evaluarea psihologică
- cunoască cine este considerat părintele psihodiagnosticului

1.1. Delimitări conceptuale

Există, într-adevăr, mai multe concepte care sunt folosite pentru a evidenția posibilitatea de măsurare şi cunoaştere a diferitelor aspecte şi componente ale vieții psihice. Am putea enumera câteva: evaluarea psihologică, psihodiagnoza, măsurarea psihologică, testarea psihologică. În literatura anglo-saxonă, pe care noi, la această disciplină, ne bazăm în cea mai mare măsură, sunt frecvent utilizați termenii: psychological testing (testarea psihologică), psychological assessment (măsurare psihologică) şi, mult mai rar, psychological evaluation (evaluare psihologică). Desigur, nu sunt diferențe foarte mari între înțelesurile şi sensurile acestor concepte. Ceea ce

Copyright © DEPARTAMENT ID 2008

este comun pentru toate este faptul că actionăm asupra individului (subiectului) cu anumite metode, mijloace, instrumente în vederea obținerii unor informații privind diferite aspecte şi componente ale psihismului. Psihodiagnoza este acțiunea de aplicare a testelor, care, la rândul lor, au obiective foarte precise. Avem nevoie să cunoaştem nivelul de dezvoltare intelectuală, aplicăm teste de inteligență, avem nevoie să cunoaştem nivelul de dezvoltare a capacităților mnezice, aplicăm teste de memorie ş.a.m.d. Deci, noțiunile de testare psihologică şi de psihodiagnostic au cam acelaşi înțeles. Măsurarea psihologică înseamnă ceva mai mult decât testarea psihologică, deoarece ea se poate face folosind nu numai testele psihologice, ci şi alte modalități de măsurare, cum este cazul diferitelor scale de măsurare a opiniilor, atitudinilor etc. Evaluarea psihologică, cel puțin după părerea noastră, implică acțiunile de testare şi de măsurare dar nu se rezumă numai la acestea, ci, în mod obligatoriu, ea necesită **acțiunile de interpretare a rezultatelor, de integrare a lor într-un set de informații care să descrie cât mai corect şi cât mai exact situația subiectului.** Deci, deşi nu în exclusivitate, testarea şi măsurarea psihologică se centrează mai mult asupra aspectelor de ordin *cantitativ*, în timp ce evaluarea psihologică se centrează mai mult asupra aspectelor de ordin *calitativ*. În general, evaluarea psihologică se finalizează cu un *Raport* în cadrul căruia regăsim incluse rezultatele examinării sau măsurării psihologice (de exemplu, coeficientul de inteligență - QI = 125) dar şi: a) judecăți apreciative privind starea subiectului (normală, anormală, deficitară etc.); b) conturarea unor cauze care au dus la apariția acelei stări; c) formularea unor recomandări privind acțiunile ce se impun în legătură cu subiectul (acțiuni educațional-recuperative, psihoterapeutice etc.). Aşadar, dintre toate noțiunile luate în discuție, evaluarea psihologică pare a avea sfera cea mai mare, deşi, trebuie să recunoaştem că, în literatura de specialitate, nu există un consens referitor la utilizarea acestui concept. De altfel, nici noi, în cadrul acestei discipline, nu vom utiliza predominant noțiunea de evaluare psihologică, deoarece interesul central al nostru va fi orientat către locul şi rolul pe care îl ocupă *testele psihologice* în activitatea psihologului.

Să vedem acum ce diferențe există între alte două noțiuni pe care le întâlnim frecvent: *psihodiagnostician* şi *psihotehnician*. Lucrurile nu sunt foarte complicate, important este să le înțelegem bine. Psihodiagnosticianul este persoana abilitată să aplice testele şi să elaboreze Raportul final de evaluare. Deci, psihodiagnosticianul este cel care face, de fapt, evaluare psihologică. El are, obligatoriu, pregătire academică (în cadrul departamentelor de psihologie), la care se adaugă o îndelungată activitate practică. Sunt unele teste extrem de complexe, care necesită mult timp pentru formare, ceea ce înseamnă că nu este suficientă doar activitatea de familiarizare din timpul anilor de studenție. Este cazul, după cum o să vedeți, a unor teste de personalitate, a unor teste de inteligență, a unor teste proiective. Psihotehnicianul este cel care învață să aplice bine şi corect anumite teste şi, bineînțeles, este vorba tot despre testele complexe. Rezultatele obținute le înaintează specialistului, psihodiagnosticianului,

Copyright © DEPARTAMENT ID 2008

abținându-se de la orice fel de prelucrare şi, mai ales, interpretare. În practică, psihotehnicianul poate fi şi o persoană care nu are pregătire în domeniul psihologiei, dar care a învățat să aplice un test sau mai multe teste respectând strict cerințele din cadrul manualelor acestor instrumente psihodiagnostice.

Să mai spunem câteva cuvinte şi în legătură cu noțiunea de *testare*. Ea a căpătat în ultimul timp o extindere foarte mare, fiind întâlnită în multe domenii, cu înțelesul de verificare. De exemplu, în domeniul tehnic, vorbim de acțiunea de testare a unor materiale, a unor motoare, a unor mijloace de transport etc. În domeniul medical-farmaceutic, vorbim de testarea unor medicamente, a unor substanțe, a unor instrumente etc. În domeniul aeronautic şi cosmic vorbim de testarea unor aeronave, a unor rachete. Nu în ultimul rând, în domeniul militar, se vorbeşte despre testarea unor arme noi, a unor noi mijloace de comunicare, de simulare etc. În domeniul psihologiei, acțiunea de testare nu se rezumă doar la cea de verificare, ci ea are un înțeles destul de diferit, după cum, de altfel, am spus mai sus. De aceea, este bine ca să fie folosite permanent sintagmele *test psihologic* şi *testare psihologică*, tocmai pentru a păstra specificul acestora şi a le feri de unele confuzii sau limitări. Trebuie, de asemenea, să facem o diferențiere clară a ceea ce înseamnă *evaluare psihologică empirică* şi *evaluare psihologică ştiințifică*. Cea empirică este de tip impresiv, „ochiometric" după cum spun unii, adică se bazează pe rezultatele cunoaşterii nemijlocite, directe, apelând la simțuri. Subiectul X este ceea ce „văd" că este sau ceea ce face. Multi părinti spun, cu convingere, că îşi cunosc foarte bine proprii copii, cadrele didactice, de asemenea, vor afirma, cu tărie, că îşi cunosc bine elevii cu care lucrează. Şi nu au dreptate ? Ar fi nedrept din partea noastră să folosim un „Nu" categoric. Problema este că un asemenea mod de a cunoaşte şi de a evalua prezintă multe riscuri, deoarece, frecvent, intervine *subiectivismul* evaluatorului, adică, în actul evaluării intervin factori ce țin de psihologia şi de personalitatea acestuia. De regulă, părinții au o imagine mult mai pozitivă despre copiii lor decât sunt aceştia în realitate. În cazul lor, factorul principal care intervine ține de afectivitatea manifestată față de copii (şi nu trebuie să fie, neapărat, criticați pentru aceasta). Evaluarea psihologică ştiințifică are, înainte de toate, un caracter *obiectiv*, deoarece ea se bazează pe utilizarea unor mijloace ştiințifice, cum este cazul testelor psihologice. Informația recoltată cu ajutorul lor trebuie să aibă prioritate în cunoaşterea şi evaluarea psihologică a subiectului. Numai cunoscând adevărata realitate psihologică putem să ne pronunțăm asupra faptului dacă aceasta este echilibrată, normală sau, dimpotrivă, ea necesită intervenții de tip educațional, recuperator sau psihoterapeutic.

Temă de reflecție/autoevaluare:
Enumerați mai multe consecințe posibile ale utilizării exclusive de către cadrul didactic a evaluării psihologice empirice.

1.2. De când a început, de fapt, evaluarea psihologică ?

Este greu de precizat exact, însă ea a început, sigur, cu foarte mulți ani înainte

Copyright © DEPARTAMENT ID 2008

de apariţia psihologiei ca ştiinţă. Evaluările psihologice erau făcute mai ales în scopuri „juridice", adică de a determina dacă cel învinuit spune sau nu adevărul referitor la o anumită faptă. Mijloacele folosite (să zicem, pe post de „teste") erau rudimentare şi, uneori, extrem de dure şi de agresive. Putem menţiona câteva: - cel învinuit trebuia să atingă foarte repede cu limba un obiect înroşit în foc. Dacă rămâneau urme, acestea constituiau dovada că respectivul este adevăratul vinovat; - acuzatul trebuia să înghită foarte repede o anumită cantitate de orez fiert şi dacă reuşea, era absolvit de vinovăţie; - mai mulţi suspecţi erau pusi in linie şi lăsaţi mai mult timp să privească un acelaşi obiect şi cel care se înroşea cel mai mult la faţă era considerat vinovat. Mai mulţi autori (Gregory, 1996; Stan, A., 2002) arată că unele forme rudimentare ale testării psihologice le putem regăsi în serviciile publice ale Chinei antice încă din anul 2200 î.e.n. Funcţionarii erau supuşi, la fiecare trei ani, unei forme de examinare, urmărindu-se gradul de potrivire al acestora cu „cerinţele" postului pe care îl ocupau. Pe măsură ce trecea timpul se producea şi o îmbunătăţire a modului de examinare, astfel că, începând cu anul 202 î.e.n., pe timpul dinastiei Han a fost introdus un examen scris la mai multe materii şi anume: legislaţie civilă, probleme militare, agricultura, fiscalitatea şi geografia. Având în vedere vremurile la care se face trimitere, trebuie să apreciem, totuşi, ingeniozitatea celor care au creat un fel de sistem de selecţie în trepte: a) examinarea preliminară, sarcina subiecţilor fiind aceea de a compune un poem în baza unei teme oferite, fiind obligaţi să stea o zi şi o noapte într-o cameră relativ mică. Procentele de reuşită se înscriau între 1 şi 7; b) cei reuşiţi în prima etapă treceau la o nouă fază, numită „district de examinare"; aici gradul de examinare şi evaluare a candidaţilor era mult mai dificil, aceştia fiind supuşi, timp de 3 zile şi 3 nopţi, la 3 sesiuni separate. Şi aici, procentul de reuşită era destul de mic şi anume, între 1 şi 10; c) a treia şi ultima fază avea loc la Pekin. 3% dintre candidaţi reuşeau să depăşească şi această fază, primind titlul de mandarin, ceea ce însemna că avea dreptul să fie ales pentru poziţia de înalt funcţionar public. Ce urmărea, de fapt, acest sistem de „selecţie profesională", deşi această sintagmă nu era folosită în acele vremuri ? In primul rând, depistarea unor „capacităţi" sau a unor „aptitudini" necesare pentru ocuparea unui post de funcţionar public. Era vorba, mai ales, despre frumuseţea scrisului şi claritatea exprimării, condiţii de bază pentru a asigura o bună comunicare. Deci, pe baza acestor însuşiri se putea face predicţii privind eficienţa în exercitarea funcţiilor publice.

Problema formelor de examinare şi evaluare psihologică utilizate înainte ca psihologia să devină ştiinţă a preocupat pe mai mulţi autori. Astfel, putem aminti exemplele oferite de către profesorul german Hofstäter, P., 1971 (apud Stan, A., 2002): a) în riturile de iniţiere ale societăţilor primitive erau folosite anumite probe prin intermediul cărora se stabilea dacă cei investigaţi, în special tinerii, erau în posesia unor capacităţi necesare pentru asumarea responsabilităţilor specifice adulţilor, cum ar fi: curajul, stăpânirea de sine, calităţile raţionamentului; b) în scrierile lui Platon, dedicate statului, se regăsesc precizări privind modul în care erau recrutaţi războinicii în

Copyright © DEPARTAMENT ID 2008

comunitatea ideală, aceştia trebuind să dovedească, mai ales, autodisciplină, curaj, incoruptibilitate.

Deşi aceste prime forme de evaluare psihologică sunt departe de cerinţele de ordin ştiinţific de care trebuie să ţină seama cei care se ocupă de selecţia profesională în zilele noastre, trebuie să apreciem faptul că ele şi-au dovedit utilitatea în acele vremuri şi, pe de altă parte, au influenţat pe cei care, ulterior, s-au ocupat în mod serios de problema mijloacelor folosite. După cum vom vedea mai târziu, testul psihologic nu este echivalent cu „punerea la încercare" a subiectului. În baza rezultatelor obţinute se fac predicţii în legătură cu comportamentul subiectului.

Temă de reflecţie/autoevaluare:
Care ar fi consecinţele dacă şi astăzi s-ar utiliza în activitatea de evaluare psihologică numai proba „punerii la încercare" a subiectului ?

1.3. Cine este considerat părintele psihodiagnosticului ?

Deşi foarte mulţi psihodiagnosticieni îl consideră pe A. Binet „părintele psihometriei", în special datorită faptului că el a construit, în 1905, împreună cu T. Simon, prima „scală metrică a inteligenţei", în mai toate manualele şi tratatele privind testarea psihologică anul de naştere al psihodiagnozei este considerat a fi 1890. În acel an, James McKeen Cattell a publicat, în revista *Mind*, articolul „Mental Tests and Measurements". Ar însemna, normal, că acest autor ar fi „părintele" testelor şi, respectiv, al psihometriei.

Dar asemenea preocupări privind paternitatea unor termeni nu prea îşi au rostul, deoarece psihometria are o istorie mult mai complexă, înainte de sfârşitul secolului al XIX-lea şi începutul secolului XX mulţi autori având, prin activitatea intensă desfăşurată, o parte de contribuţie la crearea acestui domeniu extrem de important, dar şi foarte controversat al ştiinţei psihologice. I-am putea aminti, în acest sens, pe Fr. Galton, Ernst Weber, Gustav Fechner, Herman Helmholtz ş.a. Desigur, nu-l putem uita pe Wilhelm Wundt, cel care a înfiinţat primul laborator de psihologie experimentală, la Leipzig, în 1879. El a folosit, de altfel, pentru prima dată termenul de „psihometrie", publicând articolul „Psychometrics Experiments" în revista *Brain*.

James McKeen Cattell a fost studentul lui W. Wundt la Leipzig, ocupându-se în cadrul tezei sale de doctorat, de diferenţele individuale privind timpul de reacţie. După terminarea tezei a predat la Bryn Mawr şi la Universitatea din Pennsylvania, întorcându-se apoi în Europa pentru a preda la Universitatea din Cambridge. Aici l-a întâlnit pe Fr. Galton, întâlnire ce-l va marca pentru toată perioada de după întoarcerea sa în SUA, mai întâi, la Universitatea din Pennsylvania şi, apoi, la Universitatea Columbia. Meritele sale sunt multiple: pe lângă faptul că a pus bazele mai multor publicaţii, cum ar fi *Psychological Review, Science, American Men of Science,* James McKeen Cattell a fundat şi celebra „The Psychological Corporation". Printre studenţii înscrişi la docotorat sub conducerea sa există şi câteva nume celebre astăzi: E.L. Thorndike (1898), care a avut mari contribuţii la dezvoltarea teoriilor

Copyright © DEPARTAMENT ID 2008

învățării şi, totodată, la dezvoltarea psihologiei educaționale; R.S. Woodworth (1899), care a publicat în 1938 unul dintre cele mai cunoscute şi mai influente tratate de *Psihologie experimentală* ; E.K.Strong (1911), autorul testului *Vocational Interest Blank*, rămas în uz şi astăzi, după ce a fost revizuit.

Un alt doctorand al său, deşi din nefericire mai puțin cunoscut, a fost Clark Wissler (1901), cel care, după unii autori (Gregory, 1996), a avut o influență covârşitoare asupra istoriei timpurii a testării psihologice. El a reuşit să coreleze scorurile la un test mintal aplicat unui număr de peste 300 de studenți de la Columbia University şi Barnard College cu rezultatele lor academice. Intenția lui era aceea de a demonstra că rezultatele la test pot fi folosite pentru prognozarea performanței academice, însă cercetările efectuate nu au confirmat aşteptările sale. Odată cu publicarea, în 1901, a unor asemenea rezultate descurajatoare, psihologii experimentali au renunțat la utilizarea timpului de reacție (RT) şi a discriminării senzoriale ca instrumente de măsură a inteligenței. După cum o să vedeți şi la alte discipline, timpul de reacție este un indicator al manifestărilor temperamentale şi nu al inteligenței.

Printre studenții lui W. Wundt s-au numărat şi alte nume sonore ale psihologiei universale şi anume: Charles Spearman, Victor Henri, Emil Kraepelin, E.B. Titchener, G. Stanley Hall, Lightner Witmer. Spearman este creditat ca fiind cel ce a creat conceptul psihometric de *fidelitate a testului (test reliability)*. Francezul Victor Henri a colaborat cu A. Binet, sugerând modul în care pot fi utilizate testele mintale pentru a măsura procesele mintale înalte. E. Kraepelin, de formație psihiatru, a fost primul experimentator al tehnicii asocierii verbale în calitate de test formal. Leightner Witmer, după ce şi-a luat doctoratul la Leipzig, reîntors în SUA, a devenit succesorul lui James McKeen Cattell la postul de director al Laboratorului de psihologie din cadrul Universității Pennsylvania. În 1897 el a înființat prima clinică psihologică din America, tot la Universitatea din Pennsylvania, iar în 1907 a scos revista *Psychological clinic*, în cadrul căreia a publicat articolul „Clinical Psychology". În felul acesta el a devenit „părintele" psihologiei clinice, deşi este puțin cunoscut în această calitate (McReynolds, 1987).

În afară de psihologi, cei care au contribuit foarte mult, chiar dacă nu direct, la crearea psihometriei sunt filosofii. Este vorba de o serie de lucrări apărute în secolele XVII, XVIII şi XIX, care cuprind idei ce vor influența foarte mult cercetările şi formulările din domeniul ştiințelor comportamentale (Gregory, 1996). Astfel, filosoful şi matematicianul Rènè Descartes s-a ocupat mult de problema modului în care sunt relaționate procesele mentale şi procesele fizice. John Locke, în lucrarea *An Essay Concerning Human Understanding*, îşi expune punctul de vedere conform căruia cunoştințele provin din experiență, idee pe care o regăsim şi în lucrările altor „empirici britanici": *A treatise Concerning the Principles of Human Knowledge* (George Berkeley); *A treatise on Human Nature* (David Hume); *Observations on Man, his Frame, his Duty and his Expectations* (David Hartley). Christian von Wolff a publicat două lucrări, *Psychologica empirica* (1732) şi *Psychologica rationalis* (1734), prin intermediul cărora lansează termenul „psihologie". De

Copyright © DEPARTAMENT ID 2008

asemenea, după unii autori (Cohen et *al*., 1996), el este primul care concepe psihometria ca ştiinţă.

Indiferent însă de disputele privind „paternitatea" unei noţiuni sau a alteia, putem afirma că, în perioada ultimelor două secole ale mileniului II, s-a produs, în domeniul psihologiei, **un salt extraordinar prin promovarea metodelor cantitative în ştiinţa psihologică**. Deci, fenomenele psihice, prin excelenţă subiective, pot fi măsurate şi evaluate matematic. S-a răspuns în felul acesta uneia dintre cele mai serioase acuze datorită căreia psihologiei nu i s-a acceptat mult timp statutul de ştiinţă. Nu ştim, astăzi, cât de mult au înţeles acest lucru contemporanii perioadei la care facem referire, însă noi, la mai bine de un secol de atunci, nu putem să nu ne exprimăm recunoştinţa faţă de toţi cei care au contribuit la realizarea acestei „construcţii" extrem de importante pentru evoluţia ulterioară a psihologiei: psihometria.

Temă de reflecţie/autoevaluare:
De ce este dificil să susţinem faptul că există un singur „părinte" al psihometriei sau a psihodiagnosticului ?

Întrebări de autoevaluare

1. Ce înţelegem prin termenul de psihodiagnoză ?
2. Ce înţelegem prin termenul de evaluare psihologică ?
3. Care sunt noţiunile, ce se referă la evaluarea psihologică , vehiculate mai frecvent în literatura de specialitate ?
4. Prin ce se deosebeşte un psihodiagnostician de un psihotehnician ?
5. De ce psihologul trebuie să promoveze în mod consecvent evaluarea psihologică ştiinţifică ?
6. Care sunt cele mai vechi forme ale testării psihologice ?
7. Ce putem reproşa celor care utilizau asemenea forme ?
8. Cui putem acorda „paternitatea" conceptului de psihometrie ?
9. Care sunt meritele principale ale lui A. Binet privind dezvoltarea psihometriei ?
10. Care sunt meritele principale ale lui James McKeen Cattell privind dezvoltarea psihometriei ?
11. Care sunt contribuţiile lui W. Wundt şi ale doctoranzilor săi la dezvoltarea psihodiagnosticului ?
12. Cu ce au contribuit şi unii filosofi la dezvoltarea psihometriei ?
13. Aţi putea să precizaţi care este locul şi rolul testelor în selecţia profesională ?

BIBLIOGRAFIE MINIMALĂ

ALBU, M. (2000). *Metode şi instrumente de evaluare în psihologie,* Cluj-Napoca: Argonaut.
MITROFAN, N. (2001), Psihometria şi direcţiile ei de dezvoltare la început de mileniu. În ZLATE M. (coord.), *Psihologia la răspântia mileniilor*. Iaşi: Polirom.
SCHIOPU, U. (2003), *Introducere în psihodiagnostic*. Bucureşti: Editura Pro-Humanitas.
STAN, A. (2002). *Testul psihologic.Evoluţie, construcţii, aplicaţii.* Iaşi: Polirom.
GREGORY, R.J. (1996). *Psychological testing*. *History, Principles, and Applications*. Needham

Copyright © DEPARTAMENT ID 2008

Heights: Allyn & Bacon.
COHEN, R.J. et *al*. (1996). *Psychological Testing and Assessment. An Introduction to Tests and Measurement*. Mayfield Publishing Company, Mountain View, ed. a 3-a.

Copyright © DEPARTAMENT ID 2008

Unitatea de învățare nr. 2
Contribuții la dezvoltarea psihodiagnosticului în secolul XX

Cuprins:
2. 1. Testele de inteligență
2. 2. Testele de personalitate
2.3. Testele de aptitudini
2.4. Testele de achiziții
2.5. Modele de testare clinică

Obiective:
La sfârşitul acestei unități de învățare studenții vor fi capabili să:

- cunoască principalele contribuții privind testele de inteligență
- cunoască principalele contribuții privind testele de personalitate
- cunoască principalele contribuții privind testele de aptitudini
- cunoască principalele contribuții privind testele de achiziții
- cunoască principalele contribuții privind testele folosite în clinică
- evalueze critic toate contribuțiile privind psihodiagnoza de până la sfârşitul sec. XX.

2.1. Testele de inteligență

Printre primele teste care au apărut la începutul secolului XX au fost testele de inteligență. După ce A. Binet şi colaboratorul său, T. Simon, au creat prima „scală metrică a inteligenței", ce cuprindea 30 de itemi, destinată identificării copiilor şcolari retardați mintal din Paris, s-a produs o intensificare fără precedent a interesului practicienilor pentru aceste instrumente. Ele vor fi aplicate în şcoli, închisori, tribunale pentru copii etc. Testul lui Binet a fost supus unor multiple revizuiri şi traduceri, atât în Europa, cât şi, mai ales, în America. Revizuirea făcută de L. Terman, în 1916, conferă acestui test denumirea de *Stanford-Binet Intelligence Scale* (Scala de inteligență Stanford-Binet). Testul lui Binet a fost un test individual, însă odată cu declanşarea primului război mondial a apărut cerința creerii unor teste care să poată fi aplicate pe grupuri mai mari de subiecți şi într-un timp mai scurt. Aşa au apărut *testele de grup* pentru abilitățile umane (the *Army Alpha* şi the *Army Betha*), create de un colectiv de psihologi avându-l în frunte pe Roberet Yerkes, care deținea şi funcția de preşedinte al Asociației Psihologilor Americani.

Copyright © DEPARTAMENT ID 2008

Odată cu apariția testelor de grup, a crescut interesul şi pentru alte categorii de teste, cum ar fi testele de aptitudini, testele de achiziții, testele de interese, testele de personalitate, deşi distincția dintre ele nu era întotdeauna uşor de făcut.

Cu puțin înainte de declanşarea celui de-al doilea război mondial, mai exact în 1939, David Wechsler a publicat prima versiune a *Scalelor de inteligență Wechsler*, şi anume, The Wechsler-Bellevue Intelligence Scale (W-B), moment de cotitură în psihometrie, deoarece aducea unele noutăți în ceea ce priveşte testarea inteligenței. Astfel, printre altele, spre deosebire de testul Stanford-Binet, care permitea calcularea doar a unui scor QI, acest nou instrument psihodiagnostic făcea posibilă calcularea mai multor scoruri şi, totodată, stabilirea profilului individual ca urmare a combinării abilităților. Ca o noutate absolută apare posibilitatea calculării QI performanță.

Cele două mari baterii de inteligență au fost supuse unor revizuiri repetate, fiecare dintre ele având ca obiectiv de bază îmbunătățirea caracteristicilor lor psihometrice. Ele s-au impus, în practica psihodiagnostică, ca cele mai de valoare teste, gradul de încredere acordat rezultatelor obținute cu ajutorul lor fiind maximal. De altfel, ele au fost preluate, traduse şi standardizate în foarte multe țări, inclusiv în țări din Europa.

Temă de reflecție/autoevaluare:
Care sunt asemănările şi care sunt deosebirile dintre cele două mari baterii de inteligență: Stanford-Binet şi Wechsler-Bellevue ?

2.2. Testele de personalitate

După al doilea război mondial încep să prolifereze testele de personalitate, care îşi propun să măsoare diferite trăsături considerate a fi „dispoziții relativ de durată care diferențiază un individ de altul" (Kaplan şi Saccuzzo, 1993). De altfel, primul test de personalitate a fost elaborat în timpul primului război mondial, purtând numele de *Woodworth Personal Data Sheet* (Gregory, 1996). Fiind publicat în formă finală după război, acest test devine reprezentativ pentru categoria de teste structurate de grup gen „creion-hârtie". Aceste teste erau prevăzute cu răspunsuri de tip „Adevărat/Fals" sau cu răspunsuri multiple la alegere, ceea ce făcea ca ele să poată fi aplicate pe grupuri mari de subiecți. Interesul pentru asemenea tipuri de teste a scăzut mult până în anii '30 – '40, pentru ca, după cel de-al doilea război mondial, să crească din nou.

În 1921, în Europa, mai exact în Elveția, Herman Rorschach a publicat testul ce-i poartă numele, *Testul Rorschach*, deschizând astfel calea pentru o nouă categorie de teste şi anume, *testele proiective*. Testul a fost introdus în America de către David Levy mult mai târziu, fiind primit cu multă circumspecție. Abia după ce un student al lui Levy, Sam Beck i-a investigat ştiințific proprietățile şi a comunicat rezultatele în cadrul tezei sale de doctorat, s-a produs o creştere rapidă a intersului pentru acest test şi, apoi, pentru testele proiective. Una dintre dovezi o constituie dezvoltarea, în 1953, de către Henry Murray şi Christina Morgan a testului numit *the Tematic Apperception Test* (TAT), mult mai structurat decât testul Rorschach. În

Copyright © DEPARTAMENT ID 2008

1928, A.F. Payne a propus *tehnica completării frazei*, care consta, în principiu, în a oferi subiectului un început de frază (o "rădăcină"), de exemplu: "Sunt foarte preocupat când…", acesta vând sarcina de a completa fraza respectivă.

O altă contribuţie remarcabilă legată de categoria testelor proiective a avut-o F.L. Goodenough, care a încercat să determine nu numai nivelul intelectual, ci şi interesele şi trăsăturile de personalitate ale copiilor pe baza analizei desenelor acestora. Dar o variantă mult mai bine structurată şi standardizată a testelor "de desen" a apărut în anul 1948, sub denumirea *Testul House-Tree-Person* şi legată de numele lui J. Buck.

În Europa testarea proiectivă era dominată de *Testul Szondi*, elaborate de L. Szondi, psihiatru elveţian de origine maghiară. Referitor la conţinutul testului, acesta consta din 48 de fotografii ale unor pacienţi psihiatri, împărţite în 6 seturi cuprinzând următoarele 8 tipuri: homosexual, epileptic, sadic, isteric, catatonic, paranoic, maniac şi depresiv. In concepţia autorului tulburările psihiatrice majore sunt cauzate de gene recesive. Un alt autor, S. Deri a adus acest test in SUA însă a renunţat la explicaţiile lui Szondi. După opinia lui alegerea fotografiilor s-ar datora identificării inconştiente a subiectului cu caracteristicile pacienţilor fotografiaţi.

Un moment crucial în evoluţia testelor de personalitate îl reprezintă dezvoltarea, în anul 1943, a *Inventarului Multifazic de Personalitate Minnesota (Minnesota Multiphasic Personality Inventory* – MMPI). Spre deosebire de testele de personalitate structurate, de genul testului Woodworth, autorii testului MMPI au argumentat că înţelesul răspunsurilor la test poate fi determinat prin cercetări empirice. După aproape o jumătate de secol de utilizare, MMPI va apărea într-o nouă versiune, respective, MMPI-2 (Butcher, 1989, 1990). În toată această perioadă el s-a impus ca unul dintre cele mai utilizate teste de personalitate, în legătură cu el fiind elaborate multe mii de lucrări.

Un alt test de personalitate celebru, care a fost dezvoltat în acelaşi context al răspunderii pentru cercetarea empirică, a apărut în 1957, sub denumirea *California Psychological Inventory* (CPI). Deoarece variantele originale ale testelor MMPI şi CPI se confruntau cu unele probleme psihometrice semnificative, revizuirea lor – pentru MMPI, în 1986, iar pentru CPI, în 1987 – a făcut ca aceste probleme să fie înlăturate, testele căpătând o largă utilizare nu numai în SUA, ci în mai multe ţări ale lumii. Tot cam în aceeaşi perioadă cu MMPI a apărut şi testul *16 PF Cattell (the Sixteen Personality Factor*), dezvoltat de către R.B. Cattell şi care rămâne de.a lungul timpului un bun exemplu de test de personalitate bine strucuturat şi care este bazat pe metoda analizei factoriale.

O altă categorie de teste a fost creată pentru orientarea şi consilierea persoanelor. Amintim, mai întâi, *Inventarul de interese*, care îşi avea originea în studiul lui R.L. Thorndike (1912), efectuat pe un lot de 100 de studenţi. În perioada 1919-1920 Yoakum a dezvoltat o bază de 1000 de itemi privind interesele din copilărie şi până la maturitatea timpurie (Dubois, 1970). O mare parte din aceşti itemi au fost încorporaţi în *Carnegie Interest Inventory,*

Copyright © DEPARTAMENT ID 2008

test care a fost supus, de către K.M. Cowdery, timp de doi ani, respectiv, 1926-1927, la mai multe îmbunătăţiri, cum ar fi, de exemplu: creşterea numărului de itemi, compararea răspunsurilor a 3 grupe-criteriu (medici, ingineri şi jurişti) cu grupe de control cuprinzând neprofesionişti.

Edward K. Strong a revizuit testul lui Cowdery şi timp de 36 de ani s-a ocupat de un nou instrument destinat a măsura interesele, cunoscut sub numele *Strong Vocational Interest Blank* (SVIB). Acesta va deveni unul dintre cele mai utilizate teste din toate timpurile, deşi, a avut, mai mereu, un serios competitor testul ce purta denumirea de *Kuder Preference Record*, creat în anul 1934. Ceea ce era specific pentru acest test era faptul că el compara mai mult puterea relativă a intereselor la nivel individual decât răspunsurile individuale cu răspunsurile variatelor grupe profesionale. Deci, era un test ipsativ iar cele mai recente revizuiri ale acestui instrument psihodiagnostic include variantele *Kuder Survey* şi *Kuder Occupational Interest Survey* (Zytowski, 1985).

Teme de reflecţie/autoevaluare:

1. Care este primul test de personalitate construit în timpul primului război mondial ? Prin ce se caracterizează el ?
2. Care sunt cele mai importante teste proiective construite în prima jumătate a secolului XX ?
3. Prin ce se deosebeşte testul 16PF Cattell de testele MMPI şi CPI ?
4. Care sunt testele de interese mai cunoscute ? Prin ce se caracterizează ele ?

2.3. Testele de aptitudini

Dezvoltarea testelor de aptitudini a rămas oarecum în urma celei a testelor de inteligenţă, deşi ele sunt instrumente de măsură a abilităţilor mult mai specifice şi mai delimitate. Şi aceasta mai ales din două motive: unul statistic şi altul social (Gregory, J., 1996). Problema statistică ţinea de faptul că o nouă tehnică şi anume, *analiza factorială*, era necesară frecvent pentru a stabili care dintre aptitudini erau primare şi, totodată, distincte una faţă de alta. Analiza factorială i-a permis lui L.L.Thurstone să concluzioneze că există factori specifici ai abilităţilor mintale primare, cum ar fi: înţelegerea verbală, abilitatea numerică, abilitatea spaţială, memoria asociativă, viteza perceptuală, raţionamentul general. În concepţia acestui autor, în structura aptitudinilor nu există un singur factor general, aşa cum susţinuse Spearman, ci mai mulţi, respectiv, şapte. În 1938 el construieşte una dintre primele baterii de teste pentru aptitudini multiple, intitulată *The PrimaryMental Abilities Test* (PMA). Ulterior au fost dezvoltate alte baterii de aptitudini, care au fost aplicate anual pe milioane de subiecţi din SUA.

Una dintre cele mai populare baterii de teste de aptitudini a fost realizată de către Bennet, Seashore şi Wesman (1982, 1984) sub denumirea *The Differential Aptitude Test* (DAT). Prima versiune a apărut în anul 1947, fiind dedicată, iniţial, orientării vocaţionale a elevilor din clasele VIII-XII şi, apoi,

Copyright © DEPARTAMENT ID 2008

consilierii vocaţionale a adulţilor tineri şi selecţiei angajaţilor. Referitor la conţinut, această baterie cuprinde un număr de 8 teste independente, ceea ce înseamnă că, în funcţie de scopurile diagnostice, bateria poate fi aplicată şi parţial.

O acţiune de mare amploare a fost iniţiată, în 1930, de către Departamentul muncii din SUA, constând în crearea unor teste de aptitudini necesare pentru prognozarea performanţei în muncă pentru 100 de ocupaţii specifice. Ulterior, mai exact în 1940, acest departament apelează la serviciile unor profesionişti în măsurare şi psihologie industrial-organizaţională pentru a crea o baterie de teste pentru aptitudini multiple, capabilă să măsoare ocupaţiile studiate mai înainte. Aşa a luat naştere *General Aptitude Test Battery* (GATB), prima baterie folosită pentru predicţia performanţei în muncă. Referitor la conţinut, această baterie cuprinde 8 teste „creion-hârtie" şi 4 teste-aparate. Cele 12 teste pot fi aplicate în 2 ore şi ½ şi permit calcularea scorurilor la 9 factori.

Unul dintre cele mai folosite teste de aptitudini, tip „creion-hârtie" este ASVAB (*The Armed Services Vocational Aptitude Battery*). Anual el este aplicat pe un număr de peste 2 milioane persoane şi cuprinde 10 subteste.

2.4. Testele de achiziţii

Din această categorie fac parte două grupe de teste şi anume: a) teste folosite pentru admiterea în instituţiile de învăţământ; b) teste propriu-zis de achiziţii (*achievement tests*). Din prima grupă fac parte multe teste, ele avându-şi originea în testele de inteligenţă *The Army Alpha* şi *The Army Betha*, folosite în timpul primului război mondial şi fiind destinate a măsura inteligenţa subiecţilor. Unul dintre cele mai vechi este *The College Entrance Examination Board* (CEEB). După introducerea maşinii de scorare, în 1930, aceste teste au evoluat către *College Board Tests*, în particular *The Scholastic Aptitude test, cunoscut în prezent sub numele de Scholastic Assessment Test*. Funcţiile CEEB au fost subsumate ulterior la *Educational Tests Service* (ETS), care s-a preocupat de dezvoltarea, standardizarea şi validarea unor teste folosite pentru admitere, devenite apoi foarte cunoscute, precum: *The Graduate Record Examination, The Law School Admissions Test, Peace Corps Entrance Tests*.

Testele de achiziţie au ca o trăsătură comună faptul că ele urmăresc diagnosticarea, la nivel individual, a achiziţiilor realizate de către cei incluşi în procesul de învăţământ, fie acestea cunoştinţe din diferite domenii, fie deprinderi şi capacităţi formate. Deci este vorba şi de aspectul informaţional (achiziţionarea de cunoştinţe) şi de aspectul formativ (capacitatea de a opera, de a acţiona). Toate acestea sunt teste standardizate, ceea ce înseamnă că rezultatele obţinute de un elev, de exemplu, sunt raportate la rezultatele obţinute de întreg lotul de subiecţi folosit pentru standardizare

Temă de reflecţie/autoevaluare:
Care sunt elementele de asemănare şi cele de diferenţă între testele de aptitudini şi cele de achiziţie ?

2.5. Modele de testare clinică

Unii autori au încercat să facă o evaluare a modului în care psihometria s-a implicat şi a contribuit la rezolvarea problemelor specifice unui anumit

Copyright © DEPARTAMENT ID 2008

domeniu al psihologiei aplicate. Astfel, de exemplu, S.J. Korchin şi D. Schuldberg (1981) au identificat patru modele sau orientări privind testarea psihologică în clinică. Ele reflectă diferenţele profunde privind concepţia despre natura umană, distresul psihologic şi condiţiile pentru schimbarea terapeutică efectivă:

Primul model, de altfel şi cel mai vechi este **modelul psihometric** şi el are ca element central măsurarea pe bază de teste. Scopul îl reprezintă predicţia statistică a trăsăturilor şi el reclamă ca testele utilizate să fie valide şi demne de încredere. Obiectivitatea itemilor testului trebuie abordată prin obiectivitatea examinatorului.

Al doilea model poartă denumirea de **tradiţia clinică**. În acord cu acest model evaluatorul clinic poate utiliza tehnici psihometrice, însă, spre deosebire de modelul psihometric, al cărui scop era descrierea trăsăturilor, el se concentrează asupra descrierii personalităţii, incluzând arii multiple şi niveluri de funcţionare. Evaluatorul clinic, acţionând în acord cu tradiţia clinică, pune mai mult accentul pe utilizarea raţionamentului, inferenţei şi subiectivităţii.

Al treilea model **– măsurarea comportamentului –** diferă mult de modelele anterioare, atât în concepţie, cât şi în practică. Evaluatorii comportamentului ocolesc concepţiile asupra personalităţii privind dispoziţiile sau trăsăturile. Ei resping ideea conform căreia personalitatea are legătură cu ceea ce un individ **este** sau **are** şi pun accentul mai mult pe ceea ce individul **face**. De aceea, rolul testării psihologice este minimizat. Criteriile psihometrice sunt în mare măsură incompatibile cu presupunerile comportamentale.

Al patrulea model îl reprezintă **psihologia umanistă**, ai cărei reprezentanţi se manifestă predominant negativ faţă de măsurare. C. Rogers (1942) a avut o mare influenţă în această direcţie, susţinând dezavantajele utilizării testelor. În concepţia lui clientul şi nu terapeutul trebuie să fie „diagnosticianul". Făcând o sinteză a mai multor puncte de vedere, A. Sugarman (1978) evidenţiază următoarele argumente pentru afirmaţia că măsurarea psihologică nu este umanistă: - măsurarea este reducţionistă; - măsurarea este artificială; - măsurarea nu acordă atenţie relaţiei examinator-pacient; - măsurarea judecă pacientul; - măsurarea este prea intelectuală.

Întrebări de autoevaluare:

1. Care este contribuţia lui A. Binet şi a lui T. Simon la dezvoltarea psihodiagnozei ?
2. Câţi itemi cuprinde prima „scală metrică a inteligenţei" ?
3. Care sunt primele teste de grup folosite pentru testarea abilităţilor umane ?
4. Care este contribuţia lui L. Terman privind dezvoltarea psihodiagnosticului ?
5. Dar a lui D. Wechsler ?
6. Care este cel mai vechi test de personalitate ?
7. Care sunt primele teste proiective create ?
8. Care sunt cele mai cunoscute şi cel mai mult folosite chestionare de personalitate ?
9. Care este bateria de teste creată de L.L. Thurstone ?
10. Care este specificul testelor de achiziţie ?
11. Care sunt diferenţele de bază dintre cele 4 modele de testare clinică ?

Copyright © DEPARTAMENT ID 2008

BIBLIOGRAFIE MINIMALĂ

GREGORY, R.J. (1996). *Psychological Testing. History, Principles, and Applications*. Alyyn & Bacon, Needham Heights.
HORGHIDAN, V. (1998), *Metode de psihodiagnostic*. Bucureşti: Editura didactică şi pedagpogică.
MITROFAN, N. (2001), Psihometria şi direcţiile ei de dezvoltare la început de mileniu. În M. Zlate (coord.), *Psihologia la răspântia mileniilor*. Iaşi: Polirom.
SCHIOPU, U. (2003). *Introducere în psihodiagnostic*. Bucureşti: Editura Pro-Humanitas.

LECTURI SUPLIMENTARE

ALBU, M. (2000). *Metode şi instrumente de evaluare în psihologie*. Cluj-Napoca: Argonaut.
EYSENCK, H.J. (1998). *Teste de inteligenţă*. Bucureşti: Queen.

Copyright © DEPARTAMENT ID 2008

Unitatea de învățare 3

Tendințe şi direcţii de dezvoltare a psihodiagnosticului la începtul mileniului III

Cuprins:
3.1. Teoria testelor
3.2. Construcția şi dezvoltarea unor noi teste
3.3. Computerizarea testelor
3.4. Predicții pentru viitorul apropiat

Obiective:
La sfârşitul acestei unități de învățare, studenții vor fi capabili să:
- cunoască diferențele dintre teoria clasică şi teoria răspunsului la item
- cunoască noile direcții de construcție şi dezvoltare a unor teste
- evalueze critic utilizarea computerului în testarea psihologică
- cunoască principalele predicții pentru viitorul apropiat

3.1. Teoria testelor

În planul teoriei, modelul clasic psihometric este înlocuit de un model nou şi anume, IRT (*Teoria răspunsului la item*). Făcând o sinteză a mai multor puncte de vedere, M. Albu (1998) ajunge la concluzia că această teorie are la bază următoarele postulate: 1. Comportarea unui subiect la un item al unui test poate fi explicată (sau prezisă) cu ajutorul unui set de factori, numiți *trăsături, trăsături latente* sau *abilități*. Orice construct inobservabil, presupus continuu, în privința căruia o teorie psihologică afirmă că persoanele se deosebesc între ele poate fi privit ca factor. Fiecărui factor i se asociază o variabilă cu valori numerice cuprinse între - ∞ şi + ∞, continuă, denumită *variabilă latentă*. 2. Se inferează existența unui factor numai dacă se observă că răspusnurile la itemi prin care se doreşte măsurarea constructurlui covariază (variază la fel). 3. Relația dintre performnța la item a subiecților şi fiecare dintre trăsăturile care au legătură cu performanța poate fi descrisă prin cîte o funcție crescătoare, denumită *funcție caracteristică a itemului* sau *curbă caracterisitică a itemului*. Această funcție arată cum depinde probabilitatea de a răspunde corect (sau afirmativ) la item de nivelul trăsăturii.

Cei care susțin acest model caută să evidențieze şi argumentele sau rațiunile privind importanța lui şi anume: 1. IRT poate compara teste alcătuite explicit din itemi diferiți. În consecință, ea permite comparații între diferite ocazii pentru acelaşi subiect, în raport cu care memoria pentru răspunsurile anterioare este o problemă, chiar dacă cele două teste nu au itemi comuni. Aceasta este numită „măsurarea liberă a testului" şi este importantă pentru testarea ajustată şi pentru testarea adaptativă computerizată. 2. Subiecții cu

Copyright © DEPARTAMENT ID 2008

acelaşi scor clasic pot să difere în ceea ce priveşte deprinderea măsurată, depinzând de presupunerile făcute de modelul IRT. 3. Estimarea clasică a nivelului atributului (deprinderii) sau numărul corect la un test de abilităţi nu este legat linear de forma deprinderii. În consecinţă, scala numărului corect nu este o scală de interval. 4. Estimările clasice ale dificultăţii şi discriminării, cum ar fi probabilitatea unui răspuns corect, valoarea *p* şi corelaţia *scor la item – scor total* (r_{it}) nu sunt dependente una de alta aşa cum sunt dependente de abilităţile subiectului. Astfel, de exemplu, un item a cărui valoare *p* în populaţia generală este 0,5 va avea o valoare mai scăzută printre cei cu abilităţi inferioare, dar psihometria clasică nu poate să prezică magnitudinea descreşterii, ceea ce nu este cazul IRT.

O altă teorie ce s-a impus în atenţia psihodiagnosticienilor este *teoria stărilor şi a trăsăturilor latente*, care încearcă să dea răspuns la următoarele întrebări (Albu, M., 1998): - cum putem afla dacă răspunsurile date de un subiect la un chestionar care măsoară o trăsătură de personalitate intervin sau nu efectele situaţiei de măsurare ? ; - dacă există efecte ale situaţiei, ce relaţie există între scorul obţinut la scală şi trăsătura măsurată ?; - ce relaţie există între stări şi trăsături ?

Dar ce este *starea* şi ce este *trăsătura* ? Spre deosebire de disciplinele tradiţionale ale psihologiei, care aveau în atenţie fie numai diferenţele interindividuale (cum este cazul psihologiei diferenţiale), fie diferenţele intraindividuale (cum este cazul psihologiei generale), *teoria stărilor şi trăsăturilor* recunoaşte printre atributele psihologice existenţa atât a deosebirilor dintre indivizi, stabile în timp, cât şi a schimbărilor intraindividuale. Ea consideră că fiecare atribut psihologic observat este afectat într-un anumit grad de: - caracteristici ale individului; - caracteristici ale situaţiei şi/sau influenţelor care interacţionează; - eroarea de măsură. În consecinţă, rezultatul măsurării unei variabile observate se descompune în: 1. o componentă care nu depinde de situaţie şi/sau de efectele interacţiunilor; 2. o componentă care depinde de situaţie şi/sau de efectele interacţiunilor; 3. o eroare de măsură. Prima componentă este denumită *trăsătură*, iar suma primelor două componente este denumită *stare*.

Existenţa acestor componente are implicaţii extrem de importante asupra modului în care se face evaluarea psihologică (Albu, M., 1998).În construirea şi utilizarea testelor este necesar să se cunoască nu numai fidelitatea acestora, ci şi cât de mult măsoară aceste caracteristici stabile ale persoanelor şi cât de mult sunt afectate ele de situaţia în care se face măsurarea. Când se urmăreşte măsurarea *trăsăturilor*, rezultatele măsurărilor trebuie să fie afectate cât mai puţin de efectele specifice ocaziei de măsurare. Când se evaluează *starea*, instrumentul utilizat trebuie să fie sensibil la influenţele situaţiei.

O altă tendinţă actuală în psihodiagnostic şi care se va accentua în anii următori este trecerea de la evaluarea psihometrică la evaluarea potenţialului de învăţare (Havârneanu, C., 2000). Aceste două sisteme de evaluare au fost percepute antagonist, iar explicaţia rezidă în apariţia unei noi paradigme care se opune practicii tradiţionale îndelung aplicate. Evaluarea potenţialului de învăţare oferă posibilitatea construirii şi exersării unei sarcini, urmărindu-se

Copyright © DEPARTAMENT ID 2008

evoluția subiecților în rezolvarea ei. Această metodă de evaluare, spre deosebire de psihometrie, nu se bazează pe ceea ce subiecții au învățat înainte, ci propune sarcini de învățare specifice pentru a se cunoaşte profitul pe care ei pot să-l obțină.

Pentru evaluarea cantitativă şi calitativă a potențialului de învățare se folosesc două procedee: 1. *Antrenamentul în timpul testului*, procedeu care constă în corectarea şi antrenarea subiectului imediat după o soluție greşită. După această fază urmează evaluarea, în care se oferă subiectului un număr de itemi pe care trebuie să-i rezolve fără asistență. Ceeea ce este important în această procedură este faptul că subiectul poate să-şi manifeste aptitudinea de a stăpâni şi aplica principiile învățate în timpul evaluării; 2. *Procedeul pre-test – post-test*, ce cuprinde trei faze: - faza *pre-test*, care constă în obținerea unei evaluări de bază a funcțiilor actuale şi care este asemănătoare testului psihometric uzual; - faza *învățării*, care constă în a expune subiectului condițiile ce favorizează reuşita sarcinii propuse. În această fază i se dau ajutoare ce-i permit să avanseze spre rezolvarea problemei, să învețe strategii de utilizat pentru rezolvare sau să-şi corecteze comportamentul indecvat față de problemă; - faza *post-test*, care constă în verificarea efectelor învățării.

De fapt, cele două tendințe nu se exclud, ci, dimpotrivă, ele sunt complementare în examenul psihologic (Hvârneanu, C., 2000). Testele permit evaluarea a ceea ce un subiect este capabil să facă în momentul administrării testului, iar evaluarea potențialului de învățare ne permite să estimăm dacă subiectul este capabil să învețe.

Teme de reflecție/autoevaluare:
1. Ce aduce nou, în domeniul psihodiagnosticului, teoria răspunsului la item (IRT) ?
2. Ce aduce nou, în domeniul psihodiagnosticului, teoria stărilor şi trăsăturilor ?
3. În ce constă evaluarea potențialului de învățare ?

3.2. Construcția şi dezvoltarea unor noi teste

Sute de noi teste sunt publicate în fiecare an şi rata proliferării lor va creşte, pe de o parte, din nevoia de a răspunde unor cerințe dinspre noi domenii (de exemplu, psihologia ecologică, psihologia comportamentului, psihologia clinică, psihologia familiei etc.), iar pe de altă parte, din nevoia de a înlocui testele mai vechi.

Printre testele mai noi putem aminti: *The Kaufman Assessment Battery for Children* (K-ABC), *Minnesota Multiphasic Personality Inventory*, varianta 2 (MMPI-2) pentru adulți şi varianta pentru adolescenți (MMPI-A), *Personality Inventory for Children* (PIC), *Multidimensional Aptutude Battery* (MAB) ş.a. Privitor la personalitate au fost construite mai multe teste bazate pe modelul celor 5 factori (BIG-FIVE). Deşi diferiți cercetători au folosit termeni diferiți, aceşti factori sunt (Minulescu, M., 1996): Nevrotismul–*Neuroticism;* Extraversiunea–*Extraversion*; Deschiderea la experiență–*Openness to Experience*; Agreabilitatea–*Agreeableness*; Conştiinciozitatea – *Conscientousness*. Rearanjând aceşti factori se ajunge la un acronim simplu: OCEAN (Gregory, R.J., 1996).

Copyright © DEPARTAMENT ID 2008

Plecând de la acest model, P.T. Costa (1991) şi R. McCrae (1989, 1992) au construit două teste de personalitate: 1. *The revised NEO Personality Inventory* (NEO-PI-R), ce conţine 240 de itemi. În afara celor 5 domenii majore ale personalităţii, inventarul măsoară 6 trăsături specifice, numite faţete, în cadrul fiecărui domeniu; 2. *The NEO Five-Factor inventory* (NEO-FFI), ce cuprinde 60 de itemi şi care este, de fapt, o versiune prescurtată a celuilalt.

Oricum, noile teste îşi dovedesc superioritatea fie datorită unor calităţi psihometrice superioare, fie datorită faptului că sunt mult mai specifice şi mult mai potrivite pentru diferite probleme particulare. De asemenea, testele mai noi se dosebesc fundamental de cele tradiţionale, deoarece ele sunt bazate pe concepte şi teorii moderne din diferite ramuri ale psihologiei. Este suficient să amintim, de exemplu, impactul produs asupra construirii testelor de inteligenţă de teoria lui R. Sternberg (modelul triarhic) şi de teoria lui H. Gradner (variante multiple ale inteligenţei).

De altfel, proliferarea testelor netradiţionale este determinată de două tendinţe existente în testarea psihologică: 1. testele netradiţionale reflectă creşterea răspunsului ştiinţei psihologice la nevoile din ce în ce mai mari de aplicare în practică (Haynes, S.N.,1991). Chiar şi cei mai înverşunaţi adversari ai psihometriei pot să recunoască faptul că noile teste constituie, înainte de orice, un răspuns la obiecţii şi o încercare de perfecţionare şi de creştere a „puterii" (*soundness*) testului; 2. Se fac eforturi pentru a integra testele cu alte aspecte ale psihologiei aplicate. De exemplu, mulţi psihologi, orientaţi în special spre latura comportamentală, au fost decepţionaţi de relaţia extrem de slabă dintre măsurarea clinică, testele tradiţionale şi intervenţiile pe linia tratamentului (Haynes, S.N., 1992). Cel mai de dorit ar fi nu testele ale căror rezultate au o legătură directă cu tratamentul, ci acelea care pot fi utilizate pentru măsurarea eficacităţii tratamentului.Un alt aspect al proliferării instrumentelor psihodiagnostice îl constituie construirea unor teste din categoria celor *semistandardizate*. Este vorba, de altfel, de încercarea psihologilor de a scoate psihodiagnosticul de sub controlul exclusiv al experţilor şi de a oferi şi altor utilizatori (cum este cazul părinţilor, cadrelor didactice, personalului medical, asistenţilor sociali ş.a.) unele instrumente psihodiagnostice care să-i ajute în activitatea lor. Aceste teste ar oferi informaţii cu caracter orientativ (Mitrofan, N., 1997), iar în cazul în care sunt semnale că există probleme mai serioase, trebuie să se apeleze la un psihodiagnstician expert. Şi dacă ne gândim că asemenea culegeri de teste publicate poartă pe coperţi nume celebre, cum este cel al lui H.J. Eysenck (1998), putem să înţelegem mai uşor utilitatea lor. Cu toate acestea, considerăm că sunt necesare serioase precauţii, deoarece o asemenea intenţie generoasă poate fi contrabalansată de multiple efecte negative asupra psihodiagnozei autentice, ştiinţifice.

Teme de reflecţie/autoevaluare:

1. Care sunt cele mai noi teste create ? Ce trăsături au ele faţă de cele create mai demult ?
2. De ce trebuie să manifestăm precauţie faţă de testele semistandardizate ?

Copyright © DEPARTAMENT ID 2008

3.3. Computerizarea testelor

Desigur, un prim scop al utilizării computerului în psihodiagnoză îl reprezintă scurtarea timpului de aplicare, incluzând toate etapele şi, mai ales, scorarea şi interpretarea rezultatelor. Unii autori încearcă să evidenţieze şi alte avantaje. Astfel, C. Havarneanu (2000) delimitează următoarele criterii care evidenţiază avantajele utilizării computerului în examinarea psihologică: 1. *Criteriul timp*. Faţă de probele clasice, cele computerizate aduc o condensare temporală. Aceasta nu rezultă din scurtarea timpului de solicitare, ci din modul rapid de prelucrare, afişare şi tipărire a rezultatelor. Timpul câştigat poate fi alocat întreţinerii psihologice prelungite cu subiectul aflat în situaţia de examinare. Computerul nu se interpune între examinat şi examinator, el oferind posibilitatea prelungirii sensibile a contactului uman direct, atât de necesar realizării unui psihodiagnostic competent; 2. *Criteriul mobilităţii*. Faţă de unele probe de reactivitate senzorio-motorie utilizate, subiectul poate reacţiona la stimuli în mişcare bidimensională. Posibilitatea utilizării unor stimuli perturbatori are o gamă mai largă de utilizare. Un avantaj cert este acela că există posibilitatea subiectului de a comunica interactiv cu computerul, care posedă largi distribuţii ale posibilităţilor de răspuns; 3. *Criteriul particularizării şi individualizării examenului*. În formele tradiţionale de examinare, operativitatea este scăzută din cauza timpului practic limitat care se poate aloca pentru a culege un număr mare de date şi pentru a face comparaţii rapide ale acestora. În formele de examinare computerizată se pot efectua comparaţii rapide, se pot nuanţa rezultatele, iar interacţiunea datelor poate fi analizată în permanenţă; 4. *Criteriul economic*. O particularitate deloc neglijabilă în capacitatea de investigare a unui laborator constă în posibilitatea de dotare materială. O probă de tip clasic presupune cheltuieli de achiziţionare superioare costului unui computer pe care pot fi stocate un număr nelimitat de probe psihologice.

Un alt mod de utilizare îl reprezintă *testarea computerizată adaptativă*. Diferitele seturi de întrebări ale testului sunt administrate, cu ajutorul calculatorului, la diferiţi indivizi în funcţie de „statutul" fiecăruia dintre ei faţă de trăsătura supusă măsurării. În testarea abilităţilor, de exemplu, computerul adaptează nivelul de dificultate a itemului în funcţie de răspunsul subiectului. Dacă răspunsul este incorect, este oferit un item mai uşor, iar dacă este corect, poate fi selectat un item mai dificil. Un alt exemplu: un computer poate avea o bancă de itemi pentru un test de achiziţie, aceştia prezentând diferite niveluri de dificultate. Computerul poate fi programat: 1. să nu prezinte un item crescut de dificultate dacă subiectul nu a răspuns corect la 2 itemi succesivi de un nivel de dificultate inferior; 2. să termine testarea când subiectul nu răspunde corect la 5 itemi consecutivi de un anumit nivel de dificultate.

O altă direcţie de utilizare a computerului este aceea în care el *generează sarcini ce nu pot fi prezentate prin metode tradiţionale*. Prin intermediul calculatorului va fi posibilă abordarea unei noi palete de abilităţi ce nu au figurat în obiectivele unor teste tradiţionale.

În fiecare an sunt dezvoltate tot mai multe *programe pentru scorarea testelor şi pentru producerea unor rapoarte scrise.* M. Albu (1998) enumeră, în

Copyright © DEPARTAMENT ID 2008

lucrarea sa, mult mai multe utilităţi prezente şi, mai ales, viitoare ale programelor folosite în domeniul testării psihologice: a) păstrarea rezultatelor testării psihologice, în fişiere sau în baze de date; b) calculul scorurilor la teste; asemenea programe sunt utile mai ales în cazul testelor compuse dintr-un număr mare de itemi, repartizaţi pe mai multe scale. Există şi teste care nu pot fi cotate decât cu ajutorul calculatorului (de exemplu, testul pentru capacitatea de organizare CO92); c) administrarea testelor, urmată, desigur, de calculul scorurilor; cu ajutorul calculatorului pot fi administrate teste prin care se măsoară variabile psihice sau fiziologice care nu ar putea fi investigate în cazul administrării de tip „creion-hârtie" (de exemplu, timpul de reacţie, timpul de decizie, timpul de răspuns la fiecare item al unui chestionar etc.); d) identificarea protocoalelor invalide, în care răspunsurile subiectului nu sunt conforme cu realitatea. Un procedeu utilizat în acest scop aplică teste statistice asupra succesiunilor de răspunsuri date de subiect; e) verificarea unor ipoteze referitoare la persoana examinată, pe baza comparării, prin teste statistice, a rezultatelor obţinute de aceasta la examenul psihologic cu cele ale unui eşantion extras din populaţia căreia îi aparţine subiectul; f) intervievarea subiectului; programele conţin, alături de întrebările posibile, şi un algoritm de constituire a interviului în timpul examenului psihologic, în funcţie de răspunsurile date de subiect. Interviurile administrate de calculator sunt contraindicate însă în cazul copiilor, al adulţilor cu un nivel intelectual scăzut şi al celor cu simptome psihiatrice; g) interpretarea rezultatelor la un test psihologic, care are la bază transpunerea într-un program a unui set de reguli prespecificate, referitoare la un răspuns sau la un pattern de răspunsuri (un scor la un test sau un profil psihologic), ce permite analiza, interpretarea şi evaluarea unor calităţi ale persoanelor; h) redactarea raportului psihologic; uneori sunt formulate predicţii referitoare la subiect, fie pe baza unor metode statistice (de exemplu, folosind regresia liniară), fie pe baza identificării unor legături între scorurile la test şi unele caracteristici non-test, cum sunt datele biografice. Alteori, în urma comparării profilului psihologic al subiectului cu rezultatele obţinute la aceleaşi teste de diverse grupuri de persoane (de exemplu, grupuri care diferă între ele prin profesie, prin performanţa în muncă sau prin diagnosticul psihiatric) se determină populaţia din care face parte persoana examinată; i) alegerea tratamentului (a terapiei, a programului de instruire etc.) cel mai potrivit pentru subiect; acesta reprezintă cel mai înalt nivel de implicare a calculatorului în activitatea psihologului şi este de aşteptat ca numărul programelor de acest tip să crească; j) construirea unui test psihologic.

După unii autori (Cohen, Swerdlik şi Phillips, 1996), programele pentru computer, destinate a facilita construcţia, administrarea, scorarea şi interpretarea unor teste, cum ar fi cazul testelor de achiziţie dezvoltate de cadrele didactice, vor prolifera într-un mod impresionant. Asemenea programe, având denumiri de genul „*Make a test"*, „*Create a test"*, „*The Grand Inquisitor"*, „*The First National Item Bank and Criterion-References Scoring System*", evidenţiază două avantaje majore ale testării psihologice computerizate: 1. capacitatea de a stoca itemi în „banca de itemi"; 2.

Copyright © DEPARTAMENT ID 2008

capacitatea de a individualiza testarea printr-o tehnică numită „ramificarea itemului”

Tema de reflecție/autoevaluare:
Să presupunem că cineva manifestă reticență privind utilitatea calculatorului în activitatea psihodiagnostică. Cu ce argumente l-ați putea convinge de contrariul ?

3.4. Predicții pentru viitorul apropiat

Rferitor la tendințele de dezvoltare în viitor ale psihometriei, unii autori (Kaplan, R.M., Saccuzzo, D.P., 1993) au formulat o serie de predicții:

Predicția 1: perspectivele sunt promițătoare. Acest optimism este bazat pe rolul imens pe care l-a jucat testarea în dezvoltarea şi recunoaşterea psihologiei în general şi a psihologiei profesionale, în particular. Poate că testele, aşa cum sunt cunoscute ele astăzi, vor fi retrase din scenă, dar aceasta nu înseamnă că psihometria îşi va închide porțile, ci, dimpotrivă, ea va înflori în secolul următor;

Predicția 2: proliferarea unor teste noi şi îmbunătățite va continua cu şi mai mare intensitate. Testele de inteligență, aşa cum se prezintă ele astăzi, sunt departe de a fi perfecte, cu toate revizuirile făcute. Prin urmare, rolul dominant al bateriilor de teste Stanford-Binet şi Wechsler nu este deloc sigur pentru viitor. Chiar dacă multiplele revizuiri efectuate până acum au condus la îmbunătățirea unor elemente de conținut şi la unele calități psihometrice noi, aceste teste nu diferă în esență de caracteristicile şi de concepția care a stat la baza construirii scalelor originale. Referitor la testele de personalitate, se pare că varianta recentă (MMPI-2) a Inventarului Multifazic de Personalitate Minnesota va fi testul secolului XXI, iar în ceea ce priveşte categoria testelor proiective, testul Rorschach îşi va câştiga un nou nivel de acceptanță şi respectabilitate în secolul XXI.

Predicția 3:schimbări revoluționare tip „perestroika” în testarea şcolară. Unii specialişti susțin că, în secolul ce a început, se vor utiliza mai ales testele de achiziție standardizate la nivel național, în timp ce alții resping această idee. De asemenea, se pare că, în anii următori, *testele de performanță* ar putea înlocui testele standardizate cu răspunsuri multiple la alegere. Testele de performanță reclamă ca, în loc să ofere un răspuns verbal sau să completeze o foaie de răspuns, subiectul **să facă** ceva. Astfel, elevilor li s-ar cere să scrie eseuri, să ofere răspunsuri scrise la probleme specificate sau să rezolve probleme de matematică.

Predicția 4: vor continua controversele, neînțelegerile şi schimbările. Se pare că dezacordul şi controversa reprezintă cea de-a doua natură a psihologului. Şi nu este vorba numai de testarea psihologică, ci de orice. Desigur, motivul principal al controverselor dintre psihometricieni este legat de imperfecțiunile instrumentelor psihodiagnostice; de aceea, schimbarea va fi o caracteristică constantă în acest domeniu al psihologiei aplicate

Copyright © DEPARTAMENT ID 2008

Temă de reflecţie/autoevaluare:
Analizați aceste predicții şi stabiliți dacă viitorul activității psihodiagnostice este de bun augur sau nu

Întrebări de autoevaluare:

1. Prin ce se caracterizează Teoria răspunsului la item (IRT) ?
2. Ce înseamnă evaluarea potențialului de învățare ?
3. Care sunt cele mai importante teste mai nou construite ?
4. Ce avantaje prezintă computerizarea testelor ?
5. Ce înseamnă testarea computerizată adaptativă ?
6. Enumerați câteva programe pentru testarea computerizată.
7. Ce se aşteaptă în viitor privind activitatea psihodiagnostică ?

BIBLIOGRAFIE MINIMALĂ

ALBU, M. (1998), *Construirea şi utilizarea testelor psihologice*. Cluj-Napoca: Clusium.

ALBU, M., Pitariu, H. (1993). *Proiectarea testelor de cunoştinţe şi examenul asistat de calculator*. Cluj-Napoca: Casa cărţii de ştiinţă.

HAVÂRNEANU, C. (2000). *Cunoaşterea psihologică a persoanei*. Iaşi: Polirom.

KAPLAN, R.M., SACCUZZO, D.P. (1993). *Psychological Testing. Principles, Applications, and Issues*. Pacific Grove: Brooks/Cole Publishing Company.

MINULESCU, M. (1996). *Chestionarele de personalitate în evaluarea psihologică*. Bucureşti: Garell Publishing House.

Copyright © DEPARTAMENT ID 2008

Unitatea de învățare nr. 4
Situația psihodiagnosticului din România în perioada actuală

Cuprins:
Probleme mai vechi şi mai noi cu care se confruntă psihodiagnosticienii
Legea 213/2004 şi Colegiul Psihologilor din România
Să privim viitorul cu optimism

Obiective:
La sfârşitul acestei unități de învățare, studenții vor fi capabili să:
- cunoască bine problemele cu care se confruntă psihologii ce folosesc teste;
- cunoască cerințele legate de aplicarea şi comercializarea testelor;
- cunoască prevederile Legii 213/2004 privitoare la Crearea Colegiului Psihologilor din România;
- identifice corect locul şi rolul Comisiei metodologice;
- fie familiarizat cu realizările româneşti în domeniul psihodiagnosticului.

4.1. Probleme mai vechi şi mai noi cu care se confruntă psihodiagnosticienii.

Trebuie, mai întâi, să subliniem faptul că, şi înainte de 1989, chiar dacă psihologia nu era apreciată de guvernanții de atunci, au existat preocupări ale psihologilor români pentru dezvoltarea psihodiagnosticului. Putem aminti, în acest sens, nume precum: Gh. Zapan, U. Schiopu, P. Constantinescu, G. Bontilă, I.M.Nestor, M. Roşca, T. Kulcsar, I. Holban, A. Cosmovici ş.a. Din nefericire, mai ales după 1977, aproape un sfert de secol „pedepsele" politice aplicate ştiințelor sociale, dar mai ales psihologiei, au afectat grav şi psihodiagnosticul.

În perioada postdecembristă însă, psihologia românească a fost repusă, în mare măsură, în drepturile ei firești. S-au reînființat secțiile de psihologie din cadrul universităților, Institutul de psihologie, au apărut noi lucrări de valoare, reviste, s-au organizat manifestări ştiințifice. A apărut, totodată, şi învățământul privat, aşa că anual avem mulți absolvenți în psihologie, care doresc să activeze ca specialişti în şcoli, clinici, firme, bănci, armată, transporturi etc. În aproape orice domeniu sunt şi trebuie să fie utilizate instrumente psihodiagnostice. Dar şi până la absolvire studenții folosesc teste pentru diferite lucrări, inclusiv pentru lucrarea de licență, pentru disertația de masterat. Nu mai vorbim de cei care fac tot felul de investigații, pe bază de teste, pentru elaborarea tezei de doctorat. Dar cum se prezintă testele folosite în practică în țara noastră ?

Am putea analiza situația lor în funcție de mai multe aspecte: a) starea echipamentului testelor; b) aplicarea testelor; c) comercializarea testelor.

Referitor la *starea echipamentului testelor*, trebuie să menționăm faptul că, în marea lor majoritate, testele se prezentau într-o situație

Copyright © DEPARTAMENT ID 2008

necorespunzătoare. Cele mai multe dintre teste erau vechi şi foarte vechi, folosite şi răsfolosite, multiplicate în fel şi chip, ceea ce făcea ca multe pagini ale testelor să cuprindă, pe lângă conținutul itemilor, tot felul de semne şi de pete, din cauza multiplicării în condiții tehnice discutabile. Erau multe teste incomplete, fără manual, fără nicio sursă privind istoricul, caracteristicile psihometrice, valoarea psihodiagnostică. Şi în ceea ce priveşte, să spunem, partea auxiliară a testelor, cum ar fi partea materială, foile de răspuns, grilele etc., au existat mari probleme, în sensul că acestea ori nu au existat, ori erau confecționate într-o manieră extrem de discutabilă. Desigur, între timp lucrurile s-au mai îmbunătățit. Au fost create teste noi, respectându-se riguros cerințele de ordin ştiințific, au fost reetalonate şi restandardizate teste ce erau folosite în practică, multe teste au fost computerizate etc.

Activitatea de *aplicare a testelor* este serios şi sever reglementată în țările în care grija pentru menținerea prestigiului ştiințific al psihologului este permanentă. Există coduri, standarde etc. care precizează foarte clar cine utilizează testele psihologice, cum le aplică şi cum sunt folosite rezultatele. La noi în țară, în perioada de după Revoluție, nu au existat asemenea preocupări iar consecințele negative nu au întârziat să apară. Multe teste puteau fi aplicate de oricine, fără nicio restricție. Unii psihologi au publicat chiar cărți în care au inclus informații complete privind aplicarea testelor, cum este cazul unor teste proiective. Ei au încălcat flagrant unele prevederi ale codului deontologic. Din nefericire, unele teste ajunseseră în mâna altor specialişti (ingineri, economişti, secretari etc.), fiind vorba de unele teste serioase şi pretențioase în ceea ce priveşte aplicarea şi interpretarea rezultatelor.

Comercializarea testelor ridică cel puțin două întrebări de bază: 1. de unde procură psihologul specialist testele de care are nevoie ? 2. cine se ocupă în țara noastră de comercializarea testelor ? Ideal ar fi cam aşa: un for național asigură revizuirea testelor, le reetalonează şi, prin intermediul unor firme, care trebuie să respecte strict anumite reguli de distribuire, le oferă solicitanților, dar mai ales celor care satisfac cerințele de formare şi de specializare. Până la crearea Colegiului Psihologilor din România, nimic din ce am menționat nu a existat în realitate, absolvenții de psihologie fiind obligați, dacă nu şi-au procurat din timpul facultății ceva teste prin multiplicare-copiere, să găsească o cale strict personală pentru a ajunge la unii psihologi care folosesc teste. Şi iarăşi multiplicare, cu toate consecințele sale, teste incomplete, etaloane inutile etc. În ultimul timp mai multe firme au preluat activitatea de comercializare a testelor, ele trebuind să aibă acreditarea din partea Comisiei metodologice a Colegiului.

Temă de reflecție/autoevaluare:
Cu ce probleme s-au confruntat, şi se mai confruntă încă, psihodiagnosticienii din țara noastră ?

4.2. Legea nr. 213/2004 şi Colegiul Psihologilor din România

Desigur, necesitățile apariției acestei legi nu țin numai de problemele legate de instrumentele psihodiagnostice. Ea a apărut în 2004 sub denumirea completă „*Legea nr. 213 din 27 mai 2004 privind exercitarea profesiei de*

Copyright © DEPARTAMENT ID 2008

psiholog cu drept de liberă practică, înființarea, organizarea şi funcționarea Colegiului Psihologilor din România". Această lege trebuie să fie cunoscută de orice psiholog din România, incluzându-i şi pe cei care se află în proces de formare (studenții, masteranzii). În baza ei s-a înființat Colegiul Psihologilor din România, iar în cadrul Colegiului funcționează 4 comisii aplicative, respectiv: a) Comisia de psihologie clinică, consiliere psihologică şi psihoterapie; b) Comisia de psihologie a muncii, transporturilor şi serviciilor; c) Comisia de psihologie educațională, consiliere şcolară şi vocațională; d) Comisia de psihologie pentru apărare, ordine publică şi siguranță națională. De asemenea, mai funcționează Comisia metodologică şi Comisia de deontologie şi disciplină. Mai legată de activitatea psihodiagnostică este Comisia metodologică. Ea a elaborat deja „Normele de avizare a metodelor şi tehnicilor de evaluare şi asistență psihologică". Conform art.3 (1) Comisia Metodologică a Colegiului Psihologilor din România, prin normele de avizare prezente, stabileşte standardele de calitate şi procedurile de avizare pentru metodele şi tehnicile de evaluare şi asistență psihologică. Pe de altă parte, conform art. 3(2) „Comisiile aplicative din cadrul Colegiului Psihologilor din România, prin normele de avizare specifice, atestă competențele profesionale ale psihologilor de a utiliza metode şi tehnici de evaluare şi asistență psihologică. Toate comisiile de specialitate vor avea în vedere competențele generale de utilizare a testelor psihologice precizate în documentul de față".

De o mare importanță sunt precizările referitoare la dreptul de utilizare a testelor psihologice. Cei care folosesc testele trebuie să fie în posesia unor competențe bine precizate şi acestea se obțin prin procesul de formare profesională, însemnând şi anii de studiu şi de practică din facultate, dar şi procesul de formare postuniversitară. Categoriile majore de competențe vizează următoarele aspecte etice în aplicarea testelor: a) adoptarea unei conduite profesioniste, în acord cu normele deontologice ale psihologului şi cu respectarea legilor internaționale şi naționale privind copyright-ul (legea drepturilor de autor şi a drepturilor conexe – nr. 8/1996); b) utilizarea doar a acestor teste pentru care au competența necesară; c) asumarea responsabilității pentru modul de utilizare a testelor; d) asigurarea securității pentru testele utilizate, astfel încât ele să nu-şi piardă calitățile din cauza deconspirării publice a conținutului ori a mecanismelor de cotare; e) asigurarea confidențialității rezultatelor; f) acordul scris sau în formă electronică de includere a rezultatelor în baza de date a utilizatorului probei psihologice. De asemenea, această comisie a elaborat o serie de norme referitoare şi la modul de comercializare a testelor. Iată, aşadar, că sunt create condiții pentru ca, şi în țara noastră, activitatea psihodiagnostică să intre complet în normal.

Temă de reflecție/autoevaluare:

Faceți o scurtă analiză privind importanța elaborării Legii nr. 213/2004 pentru psihologia românească.

4.3. Să privim viitorul cu optimism.

Copyright © DEPARTAMENT ID 2008

Chiar dacă în perioada postdecembristă au fost numeroase probleme cu care s-a confruntat psihologia românească, în general, şi psihodiagnosticienii, în special, încă de la începutul ei numeroşi psihologi s-au angajat serios pe direcţia reclădirii din temelii a acestei profesii. Şi rezultatele nu au întârziat să apară. Vom aminti în continuare câteva dintre cele mai importante realizări în domeniul psihodiagnosticului:

a) apariţia unor lucrări de mari proporţii şi de mare importanţă, în cadrul cărora regăsim preocupări privind îmbunătăţirea metodologiei utilizate în cercetările psihologice(de exemplu, Zlate, M., 2000; Radu, I., 1993; Neculau, A., 1996);

b) publicarea unor lucrări româneşti, după 1990, adresate bazelor teoretice ale psihometriei (de exemplu: Minulescu, M., 1996 şi 2004; Albu, M., Pitariu, H., 1993; Albu, M., 1998 şi 2000; Havîrneanu, C., 2000; Stan, A., 2001; Horghidan, V., 1997; Mitrofan, N., 1997; Mitrofan, N., Mitrofan, L., 2006; Schiopu, U., 2003; Dumitraşcu, N., 2005 ş.a.);

c) stabilirea unor legături cu specialişti din alte ţări, fiind astfel posibil accesul la lucrări şi informaţii de mare valoare din psihodiagnosticul mondial;

d) formarea unor specialişti în alte universităţi din lume; noi înşine am activat în cadrul a două universităţi americane de prestigiu, câte 6 luni. Este vorba de University of Southwestern Louisiana, unde am lucrat cu S. Hotard şi University of Texas at El Paso from El Paso, având privilegiul de a lucra împreună cu R. Whitworth;

e) etalonarea unor teste pe populaţie românească. Am aminti aici cazul testului american DENVER , care a fost etalonat în perioada 1993-1994, cu sprijinul material şi mai ales financiar al Societăţii SORZ din Olanda (director Pieter G.J.M. Hermsen) şi cu entuziasmul unor cadre didactice şi studenţi din mai multe centre universitare. Coordonarea generală a parţinut centrului universitar Bucureşti (N. Mitrofan şi G. Drilea), iar pentru celelalte centre universitare coordonarea a fost asigurată de următoarele cadre didactice; A. Munteanu (Timişoara), C. Havârneanu (Iaşi), Ş. Szamosckosy (Cluj-Napoca);

f) pătrunderea în România a unor teste noi, moderne, care sunt folosite în cadrul unor universităţi în scop de familiarizare şi de formare a psihologilor. Este cazul testelorWAIS-III; WISC-III; WISC-III^{UK}; WPPSI-R, MMPI-2; MMPI-A; Scalele McCarthy pentru copii; Scalele de dezvoltare Bayley etc.;

g) formarea unor specialişti în cadrul programelor de masterat şi doctorat.

Întrebări de autoevaluare:

1. De ce credeţi că psihologia nu era susţinută ca ştiinţă de către autorităţi înainte de 1989 ?
2. Enumeraţi câţiva psihologi care s-au ocupat de problemele psihometriei înainte de 1989 ?
3. Cum se prezenta situaţia testelor imediat după 1990 ?
4. Cum se prezintă situaţia testelor în zilele noastre ?
5. În ce măsură Colegiul Psihologilor sprijină dezvoltarea psihodiagnosticului ?
6. Ce atribuţii are Comisia metodologică ?
7. Care sunt cele mai importante realizări în domeniul psihodiagnosticului ?

BIBLIOGRAFIE MINIMALĂ

MITROFAN, N. (2001), Psihometria şi direcţiile ei de dezvoltare la început de mileniu. În ZLATE, M., *Psihologia la răspântia mileniilor*. Iaşi: Polirom.

COLEGIUL PSIHOLOGILOR DIN ROMÂNIA (2005 şi 2006), *Acte normative*.

Copyright © DEPARTAMENT ID 2008

Modulul II – Testul psihologic ca mijloc de evaluare

Unitatea de învățare 1
Testul psihologic

Cuprins:
2.1. Cum definim testul psihologic ?
2.2. „Echipamentul" testului psihologic
2.3. Funcțiile psihodiagnozei

Obiective:
La sfârşitul acestei unități de învățare studenții vor fi capabili să:
- cunoască mai multe definiții date testului psihologic
- selecteze elementele comune din mai multe definiții
- cunoască ce înseamnă „echipamentul" (Kit-ul) complet al testului
- cunoască principalele funcții ale psihodiagnozei

2.1. Cum definim testul psihologic ?

Mai întâi, facem precizarea că termenul de **test** este legat de cel de diagnostic psihic, care, la rândul lui, provine de la cuvântul grecesc **diagnosticos**, însemnând „apt de a recunoaşte". Până la un anumit punct, diagnosticul psihic este similar cu diagnosticul medical, deoarece există o multitudine de caracteristici specifice.

Referitor la definiție, vom trece în revistă mai multe definiții, oferite de autori străini (limba engleză şi franceză) şi români:

Cronbach (1966): „Testul este o procedură sau o serie de probe, construite în scopul stabilirii prezenței (sau absenței) unui aspect psihic, a particularităților de manifestare comportamentală sau a gradului de dezvoltare psihică" .

Pierre Pichot: „Testele sunt instrumente de lucru standardizate servind de stimuli pentru un comportament, care, la rândul lui, poate fi evaluat prin compararea statistică cu comportamentul altor subiecți aflați în aceeaşi situație".

P. Oléron: „Testul este un instrument fundamental al psihologiei aplicate ce se caracterizează prin posibilitatea de a examina numeroşi indivizi în situații uniformizate".

A. Rey: „Testele psihologice sunt procedee standardizate, construite astfel încât să provoace la subiecții investigați reacții înregistrabile iar materialul obținut prin intermediul acestor teste se estimează prin referință la valorile etalon".

M. Roşca (1972): „Testul este o probă sau o serie de probe, construite în scopul stabilirii prezenței (sau absenței) unui aspect psihic, a particularităților de manifestare comportamentală sau a gradului de dezvoltare psihică".

Copyright © DEPARTAMENT ID 2008

Grand dictionnaire de la psychologie (1994): „Testul este o probă, utilizată cu precădere în psihologia diferențială, care permite descrierea comportamentului unui subiect într-o situație definită precis, prin raportare la comportamentul unui grup precis de subiecți, plasați în aceeaşi situație standard. Descrierile se fac de obicei sub formă numerică".

A. Anastasi (1992): Testul este o măsură **obiectivă** şi **standardizată** a unui **eşantion de comportament**.

Să analizăm mai amănunțit ultima definiție, în cadrul căreia am subliniat 3 concepte, considerate a fi concepte-cheie. **Măsură obiectivă** înseamnă, în principal, două aspecte: a) administrarea, cotarea şi interpretarea scorurilor sunt obiective în măsura în care acestea sunt independente în raport cu judecățile subiective ale examinatorului: b) obiectivitatea testului reiese din faptul că determinarea nivelului de dificultate al unui item ori a întregului test se bazează pe proceduri empirice, obiective şi nu pe nivelul de apreciere (deci, subiectivitate) a celui ce construieşte un test. Al doilea concept, **măsură standardizată,** înseamnă, de asemenea**,** mai multe aspecte: a) standardizarea implică uniformitatea procedurii în administrarea, cotarea şi evaluarea rezultatelor. Pentru ca scorurile diferitelor persoane să fie comparabile, condițiile de testare trebuie, în mod evident, să fie aceleaşi pentru toți. În vederea asigurării acestei uniformități a condițiilor testării, constructorul testului are obligația să ofere informații şi direcții detaliate pentru administrarea fiecărui test nou. De altfel, formularea acestor direcții este o parte majoră a standardizării unui test nou, deoarece ele se referă la: - materialele folosite; - instrucțiunile orale (instructajul); - demonstrațiile preliminare; - modalitățile de răspuns la întrebările şi neclaritățile ridicate de către subiecți; - orice alte detalii privind situația de testare; b) un alt pas important în standardizare este stabilirea *normelor*, a *etalonului*, la care raportăm, după aplicarea testului, rezultatele obținute de către subiect. Ceea ce obține concret la un test un subiect constituie *scorul brut* (raw score) şi el poate să exprime numărul corect de itemi, timpul cerut pentru îndeplinirea unei sarcini, numărul erorilor sau alte modalități de măsurare potrivite conținutului testului. Acest scor nu spune nimic până nu îl raportăm la norme sau la etalon. (Cuvântul *norme* este preluat din limba engleză şi el este echivalent cuvântului *etalon* din limba română). În sfârşit, conceptul **eşantion de comportament** se referă la faptul că, în acțiunea de testare, vizăm o anumită secvență a comportamentului sau un anumit tip de comportament, cum ar fi, de exemplu: comportament inteligent, comportament creativ, comportament motric, comportament verbal etc.

Să mai poposim puțin şi asupra altei definiții dată testului psihologic de către Robert Gregory: „Testul psihologic este o procedură standardizată folosită pentru eşantionarea comportamentului şi descrierea lui cu ajutorul categoriilor sau scorurilor". Plecând de la această definiție, putem deduce faptul că majoritatea testelor prezintă următoarele caracteristici definitorii: a) procedura de standardizare; b) eşantionul de comportament; c) scoruri sau categorii; d) norme sau standarde; e) predicția comportamentului netestat. *Procedura de standardizare* este trăsătura esențială a testului psihologic. Niciun examinator nu are voie să se abată de la setul de instrucțiuni şi de

Copyright © DEPARTAMENT ID 2008

precizările privind modul de aplicare a testului existente în cadrul manualului. Acestea reprezintă o garanție privind faptul că procedurile de administrare se aplică identic de către toți. *Eşantionul de comportament* este de interes în măsura în care el permite examinatorului să facă inferențe asupra domeniului total al comportamentului. De exemplu, scopul unui test de vocabular, care cuprinde un număr determinat de cuvinte este, de fapt, de a măsura cunoaşterea generală a vocabularului de către examinat. Pe de altă parte, un bun test trebuie să aibă capacitatea de a-i permite examinatorului să prognozeze alte comportamente şi nu pe cele reflectate de itemi. Dacă, de exemplu, răspusnul „Da" la întrebarea „Beau foarte multă apă ?" se întâmplă să ajute la prognozarea depresiei, atunci această întrebare care pare a nu avea nicio legătură devine un index util al depresiei. Aplicarea testelor se finalizează cu *derivarea de scoruri* sau de *categorii.* Întru-cât se prezumă că toți oamenii posedă trăsătura sau caracteristica ce urmează a fi măsurată într-o anumită „cantitate", scopul testării psihologice este de a estima tocmai această cantitate, exprimată numeric. De asemenea, un test psihologic trebuie să posede *norme* sau *standarde*. Rezultatele obținute de grupul de normare sau eşantionul de standardizare sunt grupate pe mai multe clase, stabilindu-se performanța medie şi, totodată, ele servesc la indicarea frecvenței cu care diferite scoruri cu valoare mai mare sau mai mică sunt obținute.

Teme de reflecție-autoevaluare:

1. Care sunt elementele de asemănare şi cele de deosebire ce rezultă în urma analizării multiplelor definiții date testului psihologic ?
2. De fapt, ce este testul psihologic ?

2.2. „Echipamentul" testului psihologic

Aşa după cum am precizat în Modulul I, din nefericire**,** mult timp testele au fost folosite în practică în variante complet necorespunzătoare: fie nu se cunoştea autorul testului, fie nu exista manual al testului, fie nu existau foi de răspuns standard etc. Orice student care îşi începe studiile în domeniul psihologiei trebuie să ştie că fiecare test trebuie să aibă un „echipament" (în limba engleză i se spune „kit") complet. Ce înseamnă asta? Înainte de orice, testul este oferit, ca orice marfă, într-un fel de ambalaj (geantă, servietă, traistă, cutie de carton etc.) pe care este înscris titlul testului. În interior regăsim două categorii de componente, respectiv:

a) testul propriu-zis, adică proba sau sarcina pe care o are de rezolvat subiectul, împreună cu instructajul şi, eventual, cu câteva exemple, pentru ca subiectul să înțeleagă mai bine ce are de făcut. Sarcinile pot fi foarte diverse, în funcție de specificul testului: poate răspunde la unele întrebări, poate efectua anumite desene, poate opera cu cuvinte sau propoziții, poate construi ceva, poate opera cu un material figural etc. De cele mai multe ori testul vizează anumite aspecte comportamentale ale subiectului şi, în funcție de rezultate, oferă informații asupra altor comportamente, considerate a fi comportamente

Copyright © DEPARTAMENT ID 2008

netestate. Testul îl poate întreba pe subiect dacă are prieteni sau nu, dacă merge la petreceri sau la întâlniri, dacă-i place să stea mai mult singur etc., răspunsurile primite măsurând comportamentul introvert sau extrovert;

b) materiale auxiliare:

b_1 – manual şi, eventual manual tehnic, în care se prezintă, în mod obligatoriu, următoarele:

- ce urmăreşte să măsoare testul, deşi intenţia poate să apară din titlu;
- autorul sau autorii, reviziile efectuate, istoricul testului;
- cui se adresează (categoria socio-profesională, vârstă, sex etc.);
- caracteristicile eşantionului pe care a fost etalonat testul;
- valoarea coeficientului de fidelitate (inclusiv procedurile în baza cărora s-a calculat acest coeficient (forma test-retest, forma alternantă, forma split-half etc.);
- valoarea coeficientului de validitate (inclusiv procedurile în baza cărora s-a calculat acest coeficient; tipul de criteriu folosit);
- precizări privind timpul de rezolvare, precauţii etc.;
- instrucţiuni privind cotarea rezultatelor (puncte, bonificări, penalizări etc.);
- norme sau etaloane, la care se raportează răspunsurile individuale.

b_2 – partea materială a testului, cum este cazul a testelor-aparte, a testelor de performanţă: - asamblări de obiecte; - aranjări de imagini; - construcţii; cuburi etc.;

b_3 – foi de răspuns, construite special, pentru a fi uşor de cotat; ele poartă numele de foi standard;

b_4 – grila sau grilele de răspuns

Desigur, în cazul în care testul este computerizat el apare pe CD şi multe din componentele auxiliare ale testului sunt adaptate. Important este, însă, faptul că pentru multe teste există ambele forme şi forma fizică, să-i spunem, şi forma electronică. Există însă şi teste care n-au cum să fie computerizate, cum este cazul marilor baterii de inteligenţă (Bateria Stanford-Binet, Bateriile Wechsler etc.), mai ales datorită subtestelor de performanţă ce presupun ca subiectul să facă ceva (să construiască, să identifice în imagini, să completeze anumite lipsuri etc.).

Temă de reflecţie/autoevaluare:

Dacă dorim să cumpărăm un test ce măsoară inteligenţa şi constatăm că vânzătorul (firma ce vinde teste) are 2 asemenea teste, pe care îl preferăm ? Deci, în baza a căror criterii ?

2.3. Funcţiile psihodiagnozei

A aplica teste pe un subiect nu poate fi un simplu scop în sine. Chiar dacă o facem „din curiozitate" tot aflăm ceva, deci tot ajungem la un rezultat, ceea ce înseamnă, de cele mai multe ori, un „diagnostic". În practica psihodiagnostică, însă, testele sunt folosite de către specialişti urmărindu-se scopuri foarte precise. De aceea, autorii (U.Schiopu, 2003) vorbeşte de existenţa mai multor funcţii ale psihodiagnozei şi anume:

Copyright © DEPARTAMENT ID 2008

a) surprinderea cât mai corect şi cât mai exact posibil a trăsăturilor şi caracteristicilor psihice individuale sau, dacă avem în vedere un grup de subiecţi, evidenţierea variabilităţii psihocomportamentale. Aplicând un test de inteligenţă pe un grup de elevi (de exemplu, o clasă) putem să scoatem în evidenţă diferenţele dintre ei în ceea ce priveşte această aptitudine. Unii vor rezulta a fi foarte inteligenţi, alţii, cu o inteligenţă de nivel mediu şi alţii, cu un nivel de inteligenţă mai redus. Şi, ştiţi care-i culmea ? Nu este obligatoriu ca cei cu un nivel de inteligenţă ridicat să aibă şi cele mai bune rezultate şcolare iar cei cu un nivel de inteligenţă sub medie să aibă cele mai slabe performanţe şcolare. Revenind la funcţia de care ne ocupăm, putem spune că acest tip de psihodiagnostic este psihodiagnosticul diferenţial;

b) evidenţierea cauzei sau cauzelor care au condus la conturarea unei realităţi psihocomportamentale, mai ales în cazul unor destructurări, dezorganizări ale sistemului psihic, luat în ansamblul său sau ale unor subcomponente ale acestuia (psihodiagnoză etiologică); un rol foarte important îl au, în această direcţie, testele proiective, mai ales testele de desen, cele care obţin informaţii despre subiect nu în manieră directă, ci în una indirectă, prin ceea ce desenează sau prin ceea ce „vede” în anumite desene, poze, planşe etc. Dacă un copil îşi desenează familia, punând în centru pe tata, care apare supradimensionat, el ne „spune” indirect despre faptul că tatăl său este hiperautoritar;

c) formularea unui prognostic, anticiparea evoluţiei psihocomportamentale în anumite situaţii şi contexte acţionale şi interacţionale; de exemplu, se fac testări pentru admiterea în diferite forme şi nivele de învăţământ, se angajează personal pentru diferite posturi de muncă sau posturi de conducere etc. În baza rezultatelor la teste, desigur dacă acestea sunt bune, se oferă anumite garanţii în legătură cu integrarea optimă într-o activitate viitoare. Dacă admiterea la facultate s-ar face pe bază de teste (de exemplu, teste de aptitudini), în baza rezultatelor obţinute, am putea anticipa că cei care au obţinut scorurile cele mai mari, vor obţine şi performanţele şcolare cele mai mari;

d) sprijinirea sarcinilor pe linia consilierii şcolare şi vocaţionale; este vorba despre acţiunile iniţiate în rândul tinerilor în vederea orientării lor către formele de şcolarizare şi de pregătire profesională. Testele psihologice ajută la depistarea capacităţilor, aptitudinilor, trăsăturilor de personalitate ale tinerilor. Toate la un loc constituie „echipamentul” psihologic individual şi, în funcţie de acesta, tânărul este orientat către aceste profesii şi, mai întâi, către acele forme de pregătire, în raport cu care tânărul posedă capacităţile sau însuşirile psihice solicitate. Ca să dăm un exemplu simplu, dacă la un tânăr constatăm că sunt bine dezvoltate aptitudinile de calcul matematic, îl vom orienta către profesiile ce solicită asemnea însuşiri psihice;

e) conturarea necesităţilor privind acordarea asistenţei psihologice în forma consilierii psihologice şi psihoterapiei. De multe ori, specialiştii apelează la diferite teste psihologice (forme obiective de măsurare) pentru a preciza mai bine condiţia psihică a subiectului (poate fi vorba de anumite stări accentuate de tensiune psihică, poate fi vorba de anumite tendinţe psihopatologice etc.). Apoi, se precizează tipul de intervenţie, cum ar fi tratamentul psihoterapeutic şi

Copyright © DEPARTAMENT ID 2008

chiar psihofarmacologic. La sfârşitul programului de terapie şi asistenţă se pot iarăşi aplica teste psihologice pentru a se vedea dacă s-au produs sau nu modificări, evident, în sens pozitiv. De exemplu, se poate constata că subiectul nu mai este la fel de depresiv sau la fel de anxios aşa cum era la început;

f) evidenţierea cazurilor de abatere de la etalonul privind dezvoltarea normală psihointelectuală şi psihocomportamentală. Specialiştii în domeniul psihologiei dezvoltării au elaborat anumite etaloane privind dezvoltarea psihică. De exemplu, pentru copiii mici, de o anumită vârstă se ştie cum trebuie să se prezinte ei din punct de vedere psihologic: câte cuvinte pot să folosească, ce tipuri de acţiuni pot să facă, cum trebuie să se raporteze la cei din jur, cum trebuie să răspundă la anumiţi stimuli afectogeni etc. Unii copii pot să „ţină" pasul cu cerinţele etalonului, alţii, însă, nu şi aceasta din multiple motive: n-are cine să se ocupe de ei, nu frecventează instituţiile preşcolare, stare de sănătate precară, lipsuri materiale etc. Cu ajutorul testelor psihologice se poate stabili dacă copilul prezintă anumite rămâneri în urmă pe linia dezvoltării şi cât de grave sunt acestea. În funcţie de rezultate se trece apoi la fundamentarea ştiinţifică a unor programe corectiv-recuperative.

g) evaluarea şi validarea unor programe de învăţare şi formare profesională şi, pe de altă parte, în baza rezultatelor, fundamentarea unor noi programe, realizarea unor corecţii sau, dacă este cazul, înlocuirea totală a unor programe;

h) formarea capacităţilor de cunoaştere şi autocunoaştere. Este cunoscut faptul că atunci când se pune problema să facem evaluări referitoare la propria noastră persoană sau în raport cu alte persoane, putem să greşim din cauza subiectivismului. Astfel, putem să ne supraapreciem sau să ne subapreciem iar atunci când ne raportăm la altă persoană, de asemenea, putem să o supraapreciem sau să o subapreciem. Or, testele psihologice, în calitatea lor de mijloace obiective de evaluare, ne ajută să surprindem realitatea aşa cum este ea. Folosind mai des testarea psihologică, ne putem aştepta şi la realizarea unor corecţii privind „grilele" de evaluare pe care le folosim în evaluarea altor persoane sau pentru autoevaluare.

Temă de reflecţie/autoevaluare:
Ce credeţi că trebuie făcut dacă, în urma testării psihologice, constatăm că un copil depăşeşte cu mult nivelul de dezvoltare psihică valabil pentru nivelul lui de vârstă ?

Întrebări de autoevaluare:

1. De ce credeţi că există mai multe definiţii date testelor psihologice ?
2. Care sunt cele mai importante elemente de asemănare a multiplelor definiţii formulate ?
3. Ce înseamnă măsură obiectivă a unui eşantion de comportament ?
4. Ce înseamnă măsură standardizată a unui eşantion de comportament ?
5. În ce constă „echipamentul" complet al unui test psihologic ?
6. La ce ne foloseşte manualul testului ?
7. În raport cu ce elemnte putem stabili valoarea testului psihologic ?
8. Care sunt principalele funcţii ale psihodiagnozei ?

BIBLIOGRAFIE MINIMALĂ

Copyright © DEPARTAMENT ID 2008

1. CLINCIU, AUREL, ION. 2005) *Psihodiagnostic*. Braşov: Edit. Universităţii din Braşov.
2. ALBU, M (1998). *Construirea şi utilizarea testelor psihologice*. Cluj-Napoca: Clusium.
3. SCHIOPU, U. (2003). *Introducere în psihodiagnostic*. Bucureşti: Ed. Pro-Humanitas.

Unitatea de învăţare 2
Clasificarea testelor psihologice

Cuprins:
2.1. Clasificarea realizată de U. Schiopu
2.2. Clasificarea realizată de H. Pitariu
2.3. Cum depăşim multiplele clasificări ?

Obiective:
La sfârşitul acestei unităţi de învăţare, studenţii vor fi capabili să:
- cunoască criteriile de clasificare a testelor propuse de U. Schiopu;
- cunoască criteriile de clasificare a testelor propuse de H. Pitariu;
- stabilească în ce măsură unele criterii de clasificare sunt exclusive sau nu
- diferenţieze şi alte criterii de clasificare a testelor psihologice

Clasificarea testelor realizată de U. Schiopu

Mai întâi, trebuie să precizăm faptul că există un număr impresionant de teste psihologice. Nimeni nu ştie numărul lor exact, dar se vorbeşte de existenţa a mii de teste. În unele ţări, cum este cazul, mai ales, a SUA, psihodiagnosticul este foarte bine dezvoltat şi foarte bine reglementat. Sunt foarte multe studii efectuate pentru a se urmări calităţile psihometrice şi în ce măsură ele sunt utile în practică. Având în vedere că există extrem de multe teste, se ridică problema clasificării lor şi, legat de aceasta, a criteriilor de clasificare. Desigur, mai mulţi autori au încercat să realizeze clasificări, însă noi ne vom referi la doi autori români.

U. Schiopu (2003) reuşeşte să diferenţieze următoarele tipuri de clasificări:
a) clasificări ce se bazează pe sarcinile psihodiagnozei, adică pe structura obiectivelor ce caracterizează testele unui sistem de diagnoză;
b) clasificări ce se bazează pe diferenţele semnificative de strategie inclusă în structura de ansamblu a bateriei;
c) clasificări ce se bazează pe tipul de material administrat în teste şi pe caracteristicile acestuia;
d) clasificări ce se bazează pe diferenţe în ceea ce priveşte forma de administrare a testelor;
e) clasificări ce se bazează pe caracteristicile tipurilor de răspunsuri solicitate;
f) clasificări ce se bazează pe tipurile de răspunsuri primite.

Clasificări în funcţie de scopul urmărit. Putem diferenţia:
- teste de cunoştinţe sau teste pedagogice;
- teste de inteligenţă;
- teste de aptitudini;
- teste pentru măsurarea unor abilităţi senzorio-motorii;

Copyright © DEPARTAMENT ID 2008

- teste de atenţie;
- teste de memorie;
- teste de personalitate;
- teste de sociabilitate;
- teste de creativitate;
- teste pentru măsurarea temperamentului;
- teste de interese;

Unele dintre aceste categorii de teste, la rândul lor, sunt profilate mai precis, în funcţie de obiectivul urmărit.

Ex. 1 - testele de inteligenţă pot fi:
- teste de inteligenţă generală;
- teste de inteligenţă tehnică;
- teste de inteligenţă abstractă;
- teste de inteligenţă socială;
- teste de inteligenţă emoţională;
- teste de inteligenţă kinestezică.

Ex. 2 – testele de atenţie pot fi teste care urmăresc ca obiective:
- concentrarea atenţiei;
- stabilitatea atenţiei;
- mobilitatea atenţiei;
- distributivitatea atenţiei.

Ex. 3 – testele de memorie pot fi:
- pentru memoria cifrelor, a figurilor;
- pentru memoria topografică

Ex. 4 – testele de aptitudini pot fi:
- pentru aptitudini tehnice;
- pentru aptitudini numerice, matematice;
- pentru aptitudini artistice;
- pentru aptitudini verbale;
- pentru aptitudini pedagogice;
- pentru aptitudini organizatorice

Clasificări ce se bazează pe diferenţele semnificative de strategie inclusă în structura de ansamblu a bateriei. Acest criteriu vizează, practic, nivelul de corelaţie dintre diferite teste. În această privinţă pot să existe următoarele situaţii:

a) între două teste poate să nu existe niciun fel de legătură (deci, nivel de corelaţie = 0), ele urmărind să măsoare în mod independent caracteristici psihice total diferite;

b) între teste poate să existe o corelaţie pozitivă, cu valoarea coeficientului de corelaţie ce tinde către valoarea + 1, atunci, mai ales, când măsoară aceeaşi însuşire psihică. Se poate da, ca exemplu, nivelul înalt de corelaţie dintre două baterii de inteligenţă, cum ar fi cazul Bateriilor Stanford-Binet şi Wechsler;

c) între mai multe teste poate apare situaţia intercorelării unor factori, cum ar fi cazul factorilor de grup; de exemplu: factorul verbal, factorul

Copyright © DEPARTAMENT ID 2008

numeric, factorul spațial. Mai multe teste pot fi saturate diferit în ceea ce priveşte fiecare dintre aceşti factori.

Clasificări în funcție de materialele şi caracteristicile materialelor utilizate:

a) trebuie diferențiate, mai întâi, testele „creion-hârtie", denumire generică ce presupune răspunsul scris pe foi de răspuns, de testele- aparate, pe de o parte (de exemplu, pompa Schultz, strungul Lahy etc.) şi testele ce folosesc diferite materiale, pe de altă parte (de exemplu, cuburile Kohs, asamblarea de imagini etc.);

b) tot din perspectiva acestui criteriu, pot fi diferențiate testele verbale şi testele nonverbale, prima categorie bazându-se pe utilizarea limbajului şi pe comprehensiunea verbală. A doua categorie măsoară o serie de capacități psihice care nu sunt condiționate de limbaj şi ele pot fi aplicate celor ce prezintă unele handicapuri şi celor care provin dintr-un mediu cultural modest sau dintr-un mediu cultural străin;

Clasificări privind forma de administrare. În funcție de acest criteriu, testele pot fi: individuale sau colective. Chiar după denumire, cele individuale se aplică exclusiv unui singur subiect; de exemplu, bateria de inteligență Stanford-Binet, testul Bayley pentru copiii mici, testele proiective etc.). Ele prezintă marele avantaj că subiectul poate fi observat tot timpul cât i se aplică testul. Testele colective se aplică pe grupuri mai mari sau mai mici de subiecți. Desigur ele se aplică şi individual. Prezintă avantajul că în timp scurt pot fi testați mulți subiecți;

Clasificări în funcție de tipul de răspuns solicitat. În raport cu acest criteriu testele pot fi grupate astfel:

a) teste cu răspuns la alegere; este cazul unor teste de personalitate, la care se alege un răspuns din două posibilități, de tipul „Da", „Nu" sau se folosesc scale de răspuns, de tipul „în foarte mare măsură, în mare măsură, în oarecare măsură, în mică măsură, în foarte mică măsură"; de asemenea, este cazul testelor de cunoştințe, la ai căror itemi există mai multe răspunsuri la alegere, printre care există şi răspunsul corect;

b) teste cu un singur răspuns bun, cum este cazul, mai ales, a testelor de inteligență şi a testelor de aptitudini. Subiectul oferă ca răspuns fie un număr, fie un cuvânt, fie o figură sau o imagine etc. Răspunsul dat este corect sau nu;

c) teste cu răspuns liber, cum este cazul, mai ales, a testelor proiective, şi mai ales a celor pe bază de planşe (de exemplu, testul Rorschach, Testul T.A.T), în cadrul cărora subiectul este lăsat liber să ofere răspunsul. În cazul acestor teste, însă, dificultățile de cotare şi de interpretare a răspunsurilor sunt mult mai mari, necesitând o mare specializare din partea practicienilor;

Clasificări în funcție de formele şi tipurile de rezultate obținute. Plecând de la acest criteriu, putem diferenția două categorii de teste: a) teste sintetice, care, în urma aplicării lor, oferă un rezultat global, de genul: QI – coeficient de inteligență; QD – coeficient de dezvoltare; Q.Kre – coeficient de creativitate; Q.E. – coeficient emoțional; b) teste analitice, care, în urma aplicării lor, permit conturarea profilului psihologic. Este cazul, mai ales, a testelor de

Copyright © DEPARTAMENT ID 2008

personalitate, chestionare sau inventare de personalitate, care conţin mai multe scale.

Teme de reflecţie/autoevaluare:
1. Care sunt criteriile folosite de U. Schipou pentru clasificarea testelor ?
În ce măsură putem considera că aceste criterii sunt suficiente ?

2.2. Clasificarea realizată de H. Pitariu (1993)

Această clasificare este făcută în funcţie de elementele care dau testului calitatea de probă standardizată: conţinutul, administrarea şi cotarea.

În funcţie de *conţinut* testele pot fi împărţite în două categorii:

1. după sarcina şi 2. după modul de procesare

După *sarcina* cu care trebuie să se confrunte persoana examinată, testele se pot împărţi astfel:

a) teste verbale, care, la rândul lor, pot fi orale sau scrise (cele de tip „creion-hârtie");

b) teste nonverbale, care pot fi împărţite în două subcategorii: b_1 – cu manipulare de aparate şi piese; b_2 – administrate cu ajutorul calculatorului;

c) teste de performanţă (subiectului i se solicită să opereze la un aparat, să completeze unele figuri lacunare sau să reproducă diferite imagini cu ajutorul unor cuburi etc.).

După *modul de procesare* implicat (adică la ce i se cere persoanei să facă pentru a da un răspuns), testele pot fi împărţite în două categorii:

a) teste intelective, care, la rândul lor, pot fi împărţite în următoarele subcategorii: a_1 – teste de performanţă; a_2 – teste de aptitudini; a_3 – teste de cunoştinţe; a_4 – teste situaţionale;

b) teste nonintelective (chestionare/inventare de personalitate)

Criteriul *administrarea* testelor. După modul în care sunt administrate, ele pot fi clasificate în termeni de *eficacitate* a examenului psihologic şi în funcţie de unele exigenţe legate de *timpul de examinare*.

În funcţie de *eficacitate*, testele pot fi împărţite în două grupe: a) teste individuale; b) teste de grup sau colective;

În funcţie de *timp*, testele pot fi grupate, de asemenea, în două grupe, respectiv: a) teste de viteză- cu limită de timp; b) teste de randament, care pot fi: b_1- cu limită de timp; b_2 – cu timp nelimitat.

Tot în raport cu criteriul administrare, putem împărţi testele în două grupe: 1. teste standardizate; 2. teste nestandardizate.

Criteriul *cotarea testelor*. Tehnica de cotare a unui test poate fi obiectivă sau subiectivă. De aceea, testele pot fi clasificate în două categorii: a) teste obiective – performanţa se apreciază după o procedură fixă, impersonală, de către un simplu operator sau cu ajutorul unei maşini de cotare; b) teste proiective (completare de fraze, pete de cerneală), cele care prezintă în cotare o doză mare de subiectivism.

Copyright © DEPARTAMENT ID 2008

Teme de reflecţie/autoevaluare

1. Care sunt criteriile folosite de H. Pitariu pentru clasificarea testelor ?
2. În ce măsură aceste criterii sunt suficiente pentru clasificarea testelor ?

Care sunt elementele de asemănare şi de diferenţă dintre clasificarea realizată de U. Şchiopu şi cea realizată de H. Pitariu ?

Cum putem depăşi inconvenientul „multiplelor" clasificări ?

Clasificările realizate nu trebuie să fie privite în mod exclusivist. Interferenţele pot fi considerate şi în sens vertical. De exemplu, dacă luăm în analiză un anumit test de inteligenţă şi urmărim clasificarea făcută de U. Şchiopu vom constata că îl putem regăsi în una din categoriile de la fiecare criteriu. Astfel, în funcţie de:

a) criteriul obiective – el măsoară inteligenţa generală;
b) criteriul strategie – testul are un caracter eterogen;
c) criteriul material administrat – test verbal, de tip „creion-hârtie";
d) forma de administrare – se aplică individual;
e) răspunsuri solicitate – un singur răspuns la fiecare item;
f) tipul de rezultate – valoarea Q.I., care este un rezultat sintetic.

Mai trebuie să facem o precizare. În niciuna din clasificările făcute nu sunt luate în atenţie testele care sunt oferite nu neapărat psihologilor specialişti, ci nespecialiştilor. De aceea, în funcţie de gradul de standardizare, credem că putem diferenţia 3 categorii de teste:

a) teste standardizate, cele care pot fi aplicate numai de către psihologi;
b) teste semistandardizate, cele care sunt destinate altor categorii de beneficiari, cum ar fi: cadrele didactice, părinţii, personal medical, asistenţi sociali etc.
c) teste nestandardizate, cele care se folosesc mai mult pentru ocuparea timpului liber .

În finalul acestei discuţii privind clasificarea testelor subliniem faptul că activitatea de psihodiagnoză nu se reduce şi nu trebuie să se reducă numai la aplicarea unui test sau a unui grup de teste. Aplicarea efectivă a unor instrumente psihodiagnostice este doar o etapă în succesiunea de etape ale activităţii de psihodiagnoză care începe cu cunoaşterea subiectului prin intermediul observaţiei conduitei, a anamnezei, a unei microanchete etc.

Întrebări de autoevaluare:

1. Ce criterii foloseşte U. Schiopu în clasificarea testelor psihologice ?
2. Cum sunt împărţite testele în funcţie de criteriul obiective urmărite ?
3. Ce criterii foloseşte H. Pitariu în clasificarea testelor psihologice ?
4. Daţi un exemplu de test şi încercaţi să-l plasaţi în funcţie de cerinţele clasificării realizate de U. Şchiopu.
5. Daţi un exemplu de test şi încercaţi să-l plasaţi în funcţie de cerinţele clasificării realizate de H. Pitariu.

Copyright © DEPARTAMENT ID 2008

4. În ce măsură utilizarea calculatorului în testarea psihologică afectează clasificarea făcută de U. Schiopu ?
4. Care sunt elementele de asemănare şi de diferenţă dintre cele două clasificări ?
5. Cum pot fi depăşite limitele impuse de multiplele clasificări ?
6. Ce sunt şi ce rol joacă testele semistandardizate ?
7. Ce sunt şi ce rol joacă testele nestandardizate ?

BIBLIOGRAFIE MINIMALĂ

ALBU, M., PITARIU, H. (1993), *Proiectarea testelor de cunoştinţe şi examenul asistat de cal-Culator*. Cluj-Napoca: Casa cărţii de ştiinţă.
CLINCIU, ION, AUREL (2005), *Psihodiagnostic*. Braşov: Editura Universităţii.
SCHIOPU, URSULA (2003), *Introducere în psihodiagnostic*. Editura Pro-Humanitas.

Unitatea de învăţare 3
Testele psihologice şi Codul deontologic

Cuprins:
3.1. Standarde pentru testarea psihologică
3.2. Psihologia ca profesie şi Codul deontologic

Obiective:
La sfârşitul acestei unităţi de învăţare, studenţii vor fi capabili să:
- cunoască preocupările la nivel internaţional privind crearea standardelor pentru activitatea psihodiagnostică;
- cunoască cerinţele privind calificările psihologului care aplică anumite categorii de teste;
- interpreteze corect prevederile existente în cadrul Codului deontologic;
- cunoască atribuţiile Comisiei de deontologie din cadrul Colegiului Psihologilor din România.

3.1. Standarde pentru testarea psihologică

La nivel internaţional au existat şi există în continuare preocupări privind elaborarea unor standarde cât mai înalte pentru testarea psihologică. Asemenea preocupări pot fi grupate în următoarele categorii:

1. *Standarde înalte pentru constructorii şi utilizatorii de teste*. Trebuie, mai întâi, subliniat faptul că, aproape paralel cu publicarea primelor teste şi, prin urmare, cu naşterea psihometriei, au apărut şi primele preocupări privind protejarea activităţii psihodiagnostice. Încă din 1895 American Psychological Association (APA) a format primul său comitet ce urma să se ocupe de măsurarea mintală şi de modul în care se aplică testele noi. Un alt comitet APA a fost înfiinţat în 1906, cu misiunea de a se ocupa de problemele de standardizare. În 1923 membrii APA resping recomandările făcute de un comitet privind monitorizarea modului de utilizare a testelor de către nepsihologi. În 1954 APA publică *Technical Recommendations for Psychological Tests and Diagnostic Tests*, document care stabilea standardele pentru testarea psihologică, precum şi recomandările tehnice. În anul imediat următor, o altă organizaţie profesională – *The National Educational*

Copyright © DEPARTAMENT ID 2008

Association-, lucrând în colaborare cu *National Council on Measurement Used in Education* (cunoscut în prezent sub numele *The National Council on Measurement)*, a elaborat şi publicat *Technical Recommendations for Achievement Tests.*

De-a lungul anilor au apărut din ce în ce mai multe publicaţii ale APA în care regăsim serioase preocupări privind asigurarea unui înalt nivel de profesionalism în utilizarea testelor psihologice (Gregory, R.J., 1996). Iată unele dintre cele mai importante:

- *Standards for Educational and Psychological Tests and Manuals* (1966);
- *Automated Test Scoring and Interpretation Practices* (1966);
- *Standards for Educational and Psychological Tests* (1974);
- *Principles for the Validation and Use of Personnel Selection Procedures* (1980);
- *Speciality Guidelines for the Delivery of Services by Clinical Psychologists* (1981);
- *Speciality Guidelines for the Delivery of Services by Industrial/Organizational Psychologists* (1981);
- *Speciality Guidelines for the Delivery of Services by School Psychologists* (1981);
- *Ethical Principles of Psychologists* (1981, 1992);
- *Standards for Educational and Psychological Testing* (1985);
- *Guidelines for Computers-Based Tests and Interpretations* (1986);
- *Standards for Educational and Psychological Testing* (1993).

The National Association of School Psychologists (NASP) a adoptat, în 1984, *Principles for Professsional Ethics*, un ghid extrem de util pentru selecţia, utilizarea şi interpretarea testelor psihologice. Acesta a fost revizuit şi adoptat în 1992.

Asociaţia Psihologilor Americani a manifestat întotdeauna grijă pentru ca testele să fie aplicate la un nivel înalt ştiinţific, dar numai de către cei calificaţi pentru aceasta. Încă din 1950, un comitet al său pentru Ethical Standards for Psychology a publicat un raport, intitulat *Ethical Standards for the Distribution of Psychological Tests and Diagnostic Aids*. Printre altele, acest raport diferenţia trei niveluri de calificare pentru utilizarea testelor şi anume:

- *Nivelul A*: teste care pot fi administrate, scorate şi interpretate adecvat cu ajutorul manualului şi al unei orientări generale (de exemplu, teste de achiziţie);
- *Nivelul B*: teste care reclamă anumite cunoştinţe tehnice privind construirea şi utilizarea testelor, precum şi cunoştinţe din alte domenii ale psihologiei, precum: psihologia diferenţială, statistica psihologică, psihologia personalului, psihologia adaptării, orientarea vocaţională (de exemplu, teste de aptitudini, inventare de adaptare aplicabile populaţiei normale).
- *Nivelul C*: teste care solicită cunoştinţe de înalt nivel privind testarea şi din alte domenii de suport, la care trebuie să se adauge supervizarea din partea unor specialişti în aceste domenii (de exemplu, teste proiective, teste de inteligenţă individuale).

Acest raport a fost integrat apoi în publicaţia *APA's Ethical Standards of Psychologists (1953)* şi citat în *Standards for Educational and Psychological Tests and Manuals* (1966).

Copyright © DEPARTAMENT ID 2008

În *Ethical Principles of Psychologists* (1992) sunt făcute precizări şi mai clare privind calificările utilizatorilor de teste. Astfel, *Principiul A* afirmă că psihologii „oferă numai acele servicii şi utilizează numai acele tehnici pentru care sunt calificaţi prin educaţie, formare sau experienţă. *Standardul etic 2.02* adaugă: „psihologii se abţin de la utilizarea necorespunzătoare a tehnicilor de măsurare şi, totodată, iau măsuri pentru a preveni ca alţii să utilizeze defectuos informaţia pe care o oferă aceste tehnici".

Cu toate măsurile luate s-a constatat că, în practică, se întâlnesc frecvente abateri. De aceea, APA, împreună cu The American Research Association, The National Council on Measurement in Education şi cu foarte multe edituri ce publicau teste, a înfiinţat The Joint Committee on Testing Practice (JCTP). Acesta a adoptat, în 1988, *The Code of Fair Testing Practices in Education*, care descrie obligaţiile celor care construiesc teste, precum şi ale celor ce le aplică, în patru arii: a) dezvoltarea/selectarea testelor; b) interpretarea scorurilor; c) strădania pentru corectitudine; d) informarea subiecţilor.

2. *Creşterea obiectivităţii în aplicarea şi interpretarea testelor*. Deşi în literatura psihodiagnostică apar lucrări favorabile testelor proiective, practicienii tind tot mai mult să se orienteze către testele obiective şi, în special către chestionare şi inventare de personalitate, având în frunte testul Minnesota Multiphasic Personality Inventory, varianta revizuită MMPI-2. Un indicator al acestei tendinţe îl reprezintă creşterea numărului de referinţe favorabile acestui test faţă de alte teste proiective şi , mai ales, faţă de testul Rorschach, aşa cum apare în The Mental Measurement Yearbook şi în alte surse (de exemplu, la Polyson, Peterson şi Marshall, 1986).

O altă tendinţă către creşterea obiectivităţii se manifestă în continuarea cercetărilor privind testarea psihologică. Deşi există mii de studii publicate în legătură cu testul MMPI, respectiv, MMPI-2, interesul pentru investigare şi cercetare nu numai că nu încetează, ci, dimpotrivă, se accentuează de la un an la altul şi va creşte cu siguranţă în viitor. În acest fel se prefigurează o creştere substanţială a orientării practicienilor către testele de personalitate, în detrimentul unor teste proiective.

Temă de reflecţie/autoevaluare:
Care ar fi consecinţele absenţei preocupărilor privind elaborarea unor standarde pentru activitatea psihodiagnostică ?

3.2. Psihologia ca profesie şi Codul deontologic.

Psihologia este o profesie foarte căutată, dovadă fiind, înainte de orice, numărul mare de candidaţi ce se prezintă la examenul de admitere la facultăţile de profil. După absolvire, psihologul poate lucra în diverse domenii, cum ar fi, instituţii şcolare, firme, instituţii militare, organizaţii guvernamentale sau neguvernamentale, instituţii medicale, în cadrul unor cabinete particulare etc. Mai peste tot el apelează la testul psihologic pentru a evalua subiectul. Această acţiune de evaluare este de mare importanţă şi ea ridică serioase probleme privind responsabilitatea specialistului. În baza rezultatelor la teste se stabilesc diagnostice şi se iau decizii privind persoana subiectului. Dacă acestea nu sunt corecte şi

Copyright © DEPARTAMENT ID 2008

concordante cu realitatea, cel care suferă este subiectul şi nimeni nu are voie să se „joace" cu soarta acestuia. În urma evaluării psihologice se poate determina dacă subiectul este apt sau nu pentru o anumită slujbă, dacă prezintă un anumit nivel de anormalitate psihică, dacă este sau nu în posesia unor capacităţi şi aptitudini etc. Dar în cazul în care diagnosticul este greşit ce se întâmplă cu subiectul ? Haideţi să ne gândim ce se întâmplă cu un copil preşcolar care, în urma testării psihologice, este orientat către formele învăţământului special, deşi, ulterior se poate constata că acest copil prezintă numai nişte rămâneri în urmă pe linia dezvoltării psihice, uşor recuperabile. Deci utilizarea testelor psihologice presupune mari răspunderi, de ordin profesional, de ordin moral şi chiar de ordin juridic.Aşa se explică de ce au existat şi există în continuare preocupări pentru elaborarea unor standarde şi, respectiv, a unor coduri deontologice.

În continuare noi vom face precizări mai ales în baza Codului deontologic al Asociaţiei Psihologilor Americani (APA), intitulat *Ethical Principles of Psychologists*. (v. şi Anastasi, A., 1996). Şi din acesta vom menţiona în special ceea ce ţine de Tehnicile de măsurare (Assessment Techiques). Vom lua în atenţie 3 categorii de probleme:

a) *legate de construcţia şi dezvoltarea testelor*:

- cei care construiesc teste şi le pun în circuit au obligaţia de a evita graba şi superficialitatea lansându-le în mod prematur pentru uzul general. Iar atunci când un test este distribuit numai în scopuri de cercetare, această condiţie trebuie specificată foarte clar;
- manualul trebuie să ofere informaţii cât mai complete şi mai corecte asupra testului, evitându-se tendinţele publicitare care să pună testul într-o lumină favorabilă;
- testele nu trebuie să fie publicate în ziare, magazine, cărţi populare, fie în scopuri descriptive, fie în scop de autoevaluare;
- testarea prin poştă sau telefonic este o practică contraindicată, deoarece, pe lângă alte riscuri, este contraindicat a se interpreta rezultatele la test în absenţa altor informaţii referitoare la subiectul examinat;
- cumpărarea, procurarea testelor este accesibilă în mod strict numai psihologilor calificaţi. Fiecare catalog trebuie să cuprindă, în mod obligatoriu, cerinţele ce trebuie îndeplinite de către solicitant sau cumpărător. De exemplu, un student care are nevoie de un test particular pentru sarcini de învăţare sau pentru cercetare trebuie să prezinte o comandă de cumpărare contrasemnată de către profesorul lor, care îşi asumă responsabilitatea pentru corecta utilizare a testului.

b) *legate de protecţia subiectului*; un factor deosebit de relevant îl constituie scopul urmărit de testarea psihologică, care poate fi: a) consilierea indivduală, cunoaşterea şi autocunoaşterea unor probleme şi aspecte psihice individuale; b) deciziile instituţionale privind selecţia şi clasificarea subiecţilor; c) cercetarea psihologică. Este evident faptul că, în raport cu scopul urmărit, subiectul se raportează diferit la testare şi la examinator. Dacă un subiect se prezintă din proprie iniţiativă la un cabinet particular, solicitând ajutor pentru anumite probleme psihice cu care se confruntă, ei vor fi mult mai cooperanţi decât subiecţii ce sunt solicitaţi instituţional să accepte acţiunea de testare. Indiferent însă de scopul urmărit de acţiunea de testare, apare obligaţia ca subiectul să fie informat (consimţământul

Copyright © DEPARTAMENT ID 2008

informat) asupra: ce se urmăreşte prin acţiunea de testare, tipul de informaţii solicitate, modul de utilizare a rezultatelor.

Consimţământul informat trebuie să fie obţinut de la subiect personal sau de la reprezentantul său legal, existând şi unele excepţii şi anume: - când testarea fără consimţământ este reglementată de lege sau de unele dispoziţii guvernamentale (de exemplu, programe de testare la nivel naţional); - când testarea apare ca o componentă a activităţii şcolare (de exemplu, testări în scop de evaluare sau de validare a unor programe educaţionale la nivel instituţional şcolar; - când consimţământul apare în mod implicit (de exemplu, testările în vederea admiterii în unele instituţii educaţionale sau angajării de personal.

c) *legate de confidenţialitate*; aici se pune întrebarea cine are acces la rezultatele obţinute în urma aplicării testelor. Evident, în primul rând, examinatorul, cel care a aplicat testul. Apoi subiectul are dreptul de a avea acces la rapoartele testării, deşi, în cazul minorilor, şi părinţii au dreptul la asemenea informaţii. În cazul în care apar anumite conflicte între dreptul copilului de a fi protejat şi dreptul părintelui de a avea acces la rezultatele testării, mai ales în cazul în care relaţiile părinţi-copii pot fi principala cauză a unor probleme psihice cu care se confruntă copilul. Desigur, dacă este vorba despre un drept el trebuie respectat, dar, în cazurile conflictuale foarte important este modul în care se face comunicarea unor informaţii în aşa fel încât aceasta să nu se soldeze cu efecte negative asupra copilului.

O altă problemă legată de păstrarea confidenţialităţii este cea privind accesul la rezultatele testării psihologice a celei de-a treia persoană, alta decât examinatul (sau părintele minorului) şi examinatorul. În acest caz se aplică principiul general, conform căruia informaţiile nu pot fi oferite celei de-a treia persoană fără consimţământul informat al subiectului.

Aşa după cum am mai menţionat, în cadrul Consiliului Psihologilor din România activează, pe lângă Comisia metodologică şi comisiile aplicative, Comisia de deontologie. Aceasta a elaborat „Codul deontologic al profesiei de psiholog cu drept de liberă practică" şi „Codul de procedură disciplinară". Codul deontologic cuprinde un set de principii, standarde etice generale, standarde specifice şi dispoziţii finale. Legat de problematica testării psihologice, ne interesează, mai ales, standardele de competenţă, standardele cu privire la relaţiile umane, standardele de confidenţialitate, standardele de înregistrare, prelucrare şi păstrare a datelor (din cadrul standardelor generale) şi capitolul „Evaluare şi diagnoză" (din cadrul standardelor specifice). Ceea ce este foarte important legat de conţinutul acestui cod este precizarea şi clarificarea unor noţiuni şi a unor sintagme. Astfel, ca să dăm câteva exemple: 1. *datele de evaluare/diagnoză* pot fi scoruri brute şi standardizate, răspunsurile subiectului la stimuli sau la întrebările la test, notele, înregistrările şi consemnările psihologului, declaraţiile şi comportamentul clientului în timpul examinării. Psihologii vor oferi datele obţinute, sub formă de rezultate clientului şi, dacă este cazul, unor terţi numai cu consimţământul clientului, sau fără acordul acestuia în condiţiile prevăzute de lege; 2. *materialele de evaluare/diagnoză* cuprind manualul instrumentului, instrumentul propriu-zis, protocoale, întrebările sau stimulii utilizaţi, alte fişe sau formulare necesare şi nu includ *datele de evaluare/diagnoză*; 3. psihologii se vor

Copyright © DEPARTAMENT ID 2008

asigura că în procesul de obținere a *consimțământului informat* următoarele puncte au fost înțelese: *scopul şi natura activității; responsabilitățile mutuale; beneficiile şi riscurile; alternativele; circumstanțele unei încetări a acțiunii; opțiunea de a refuza sau de a se retrage în orice moment, fără a suferi vreun prejudiciu; perioada de timp în care e valabil consimțământul; modul în care se poate retrage consimțământul dacă se doreşte acest lucru*; 4. psihologii vor informa persoanele fără *capacitate deplină* de a-şi da consimțământul şi persoanele pentru care testarea este cerută de reglementările legislative, cu privire la natura şi scopul serviciilor de evaluare propuse, folosind un limbaj uşor de înțeles pentru persoana care urmează să fie evaluată; 5. psihologii care folosesc serviciile unui traducător vor cere consimțământul clientului pentru a folosi serviciile acelui traducător, se vor asigura că se va menține confidențialitatea rezultatelor, securitatea instrumentelor, inclusiv a documentelor de evaluare/diagnoză.

Orice psiholog format sau aflat în procesul de formare trebuie să cunoască foarte bine conținutul celor două coduri, ca, de altfel şi celelalte acte normative aprobate de Colegiul Psihologilor din România.

Temă de reflecție/ autoevaluare

Încercați să listați cât mai multe consecințe negative posibile ale nonexistenței şi nerespectării prevederilor unui Cod deontologic.

Întrebări de autoevaluare:

1. Enumerați câteva tipuri de preocupări existente la nivel internațional privind elaborarea de standarde pentru testarea psihologică,
2. Ce credeți că înseamnă *standarde* ?
3. De ce se face diferențierea între 3 nivele de calificare pentru obținerea dreptului de a aplica anumite teste ?
4. Care este explicația faptului că se prognozează o creştere a preferinței practicienilor pentru testele obiective (în special, testele de personalitate), în detrimentul testelor proiective ?
5. Enumerați câteva cerințe etice privind construcția şi dezvoltarea testelor.
6. Enumerați câteva cerințe privind protecția subiectului, a celui examinat.
7. Ce tipuri de coduri a elaborat Comisia de deontologie şi disciplină a Consiliului Psihologilor din România ?

BIBLIOGRAFIE MINIMALĂ

ANASTASI, ANA. (1996). *Psychological Testing* (7th.ed.). New York: Macmillan.
COLEGIUL PSIHOLOGILOR DIN ROMÂNIA (2006), *Acte normative*. Ediție adăugită.
MITROFAN, NICOLAE (2001), Psihometria şi direcțiile ei de dezvoltare la început de mileniu. În M. Zlate (coord.), *Psihologia la răspântia mileniilor*. Iaşi: Polirom.
CLINCIU, ION, AUREL (2005), *Psihodiagnostic*. Braşov: Editura Universității.

Copyright © DEPARTAMENT ID 2008

Modulul III – Caracteristicile psihometrice ale testelor

Unitatea de învățare 1

Ce sunt caracteristicile psihometrice ale testelor ?

Cuprins:
1.1. Precizări conceptuale
1.2. Fidelitatea şi validitatea, caractersitci psihometrice de bază
1.3. Alte caracteristici psihometrice

Obiective:
La sfârşitul acestei unități de învățare, studenții vor fi capabili să:
- cunoască ce sunt caracteristicile psihometrice ale testelor;
- evalueze puterea diagnostică a testelor, plecând de la caracteristicile lor psihometrice;
- cunoască obligațiile constructorilor de teste privind caracteristicile psihometrice.

1.1. Precizări conceptuale.

Testele psihologice pot fi asemănate, până la un punct, cu alte instrumente de măsură, care pot fi evaluate în ceea ce priveşte calitatea şi gradul lor de precizie. Marele psiholog Alfred Binet ne avertiza pe noi psihologii să nu considerăm testul ca pe un simplu cântar, pe care dacă te urci, afli imediat ce greutate ai.. Va trebui tot timpul să ținem minte acest avertisment. Dar pentru discuția noastră, putem să facem această comparație. Există cântare de foarte mare precizie şi de foarte bună calitate, există cântare de proastă calitate şi cu un grad de precizie îndoielnic. Aşa şi testele. Unele măsoară cu mare exactitate anumite capacități şi însuşiri psihice, altele, însă, nu sunt foarte precise, nivelul de credibilitate acordat rezultatelor obținute fiind mai redus. Un test foarte bun şi foarte precis este un test *sound*, însemnând puternic, solid şi prezintă un foarte înalt grad de credibilitate. Revenind la comparația cu cântarul, acesta este făcut pentru a măsura greutatea cu el, deci nu poate fi folosit pentru a măsura altceva, înălțimea, de exemplu. Şi dacă este construit să măsoare greutatea, atunci trebuie să-şi facă foarte bine "datoria". Aşa şi cu testul psihologic. El este construit pentru a măsura o anumită însuşire psihică şi, este de aşteptat ca să o facă foarte bine. Rezultatele obținute trebuie să reflecte cât mai bine realitatea. Făcând toate combinațiile posibile, cel puțin teoretic, putem întâlni următoarele situații: a) testul măsoară ceea ce şi-a propus să măsoare şi măsoară foarte bine (soundness la nivel maximal); b) testul nu măsoară ceea ce şi-a propus să măsoare şi măsoară prost ceea ce măsoară de fapt; c) testul măsoară ceea ce şi-a propus să măsoare, dar măsoară prost; d) testul nu măsoară ceea ce şi-a propus să măsoare, dar măsoară bine

Copyright © DEPARTAMENT ID 2008

ceea ce măsoară de fapt. Cea mai bună situaţie este cea de tip a. Pentru a se menţine în această categorie, testul trebuie să "răspundă" pozitiv la întrebarea privind caracteristicile sale psihometrice: "Testul este *fidel* şi *valid*"?

Definiţie: *caracteristicile psihometrice sunt acele însuşiri ale testelor care asigură un înalt grad de credibilitate rezultatelor obţinute în urma utilizării lor în activitatea psihodiagnostică*. Cele mai importante sunt *fidelitatea* şi *validitatea*. Şi foarte important este faptul că aceste însuşiri sunt măsurabile şi evaluabile. După cum o să vedem, ambele caracteristici psihometrice pot fi măsurate şi exprimate în valori numerice: 0.50, 0.65 etc.

Pentru constructorii de teste apare obligaţia de a include, în cadrul manualului, informaţii privind valoarea celor două caracteristici psihometrice. Şi atunci este clar că, în cazul în care avem de ales între două teste care măsoară acelaşi lucru (aceeaşi însuşire psihică), dar care prezintă valori diferite pentru cele două caracteristici psihometrice, îl vom prefera pe cel care are valorile cele mai mari. Acesta este mult mai sigur în acţiunea de măsurare psihologică. Şi, de aici, mai apare o obligaţie pentru cei care construiesc teste şi anume, ei trebuie să acorde o mare atenţie respectării cerinţelor privind cele două caracteristici psihometrice încă din fazele de început ale activităţii de elaborare şi de construire a testelor. Realizaţi care sunt implicaţiile a ce spunem asupra "pieţei" psihodiagnostice ? Fiind preferate testele care stau foarte bine la cerinţele privind fidelitatea şi validitatea, cele care nu corespund vor ieşi imediat din competiţie. Şi, într-adevăr, în domeniul psihodiagnozei, mai ales a construirii de noi instrumente psihodiagnostice, trebuie să existe o serioasă competiţie.

Temă de reflecţie/autoevaluare:
Încercaţi să găsiţi cât mai multe exemple de situaţii în care s-ar folosi teste cu slabe caracteristici psihometrice şi evidenţiaţi consecinţele posibile.

1. 2. Fidelitatea şi validitatea

Definiţii: a) Fidelitatea înseamnă gradul în care testul manifestă constanţă în măsurare, adică dacă el este capabil să ofere o aceeaşi informaţie referitoare la subiectul testat în cazul în care acesta este supus unei noi acţiuni de testare folosind acelaşi test; b) Validitatea se referă la capacitatea testului de a măsura întotdeauna ceea ce şi-a propus să măsoare. Acum să explicăm puţin. În cazul fidelităţii, dacă un subiect a obţinut, de exemplu, într-o zi, în urma aplicării unui test de inteligenţă, un Q.I. = 100 şi, după câteva zile, în urma reaplicării aceluiaşi test, ar obţine un Q.I. = 70 ar însemna că, în primul moment el ar fi diagnosticat ca fiind normal, în ceea ce priveşte dezvoltarea psihointelectuală, iar în momentul 2, ar fi considerat a avea deficienţe mentale. Cine ar greşi în acest caz ? Evident că testul folosit, care nu dovedeşte fidelitate. Realizaţi ce consecinţe grave ar avea utilizarea unui asemenea test ? În limba engleză se foloseşte cuvântul "reliability" pentru a desemna fidelitatea testului, însemnând, mai ales, "demn de încredere".

Pentru ambele caracteristici psihometrice se poate calcula valoarea unui coeficient, de aceea, putem vorbi despre coeficientul de fidelitate (r_{tt}) şi coeficientul de validitate (r_{xy}). În cazul coeficientului de fidelitate, regăsim dublată litera "t", ceea ce înseamnă că este vorba de aplicarea în două momente a aceluiaşi test (este vorba de un singur instrument), iar în cazul coeficientului de validitate, avem două litere, "x" şi "y", ceea ce înseamnă că sunt folosite două instrumente de măsură. Este vorba despre testul aplicat în momentul 1 şi

Copyright © DEPARTAMENT ID 2008

alt instrument de măsurare folosit în momentul 2, care este, de regulă, la mai mare distanţă în timp folosit pentru a măsura comportamentul anticipat în baza rezultatelor din momentul 1.

Litera "r" înseamnă coeficient de corelaţie şi, deşi el poate fi calculat în mai multe moduri, cea mai utilizată procedură este cea a lui Pearson (Pearson Product-Moment Correlation Coefficient – Coeficientul de corelaţie "moment-produs" al lui Pearson). Se foloseşte următoarea formulă:

$$r_{xy} = \frac{\Sigma_{xy}}{N(SD_x)\,(SD_y)}$$

în care: x = deviaţia unei performanţe de la medie în proba x
y = deviaţia peformanţei (corespunzătoare lui x) de la medie în proba y

Revenind la conceptul de *fidelitate*, trebuie să menţionăm faptul că el a cunoscut diverse accepţiuni de-a lungul timpului. Astfel, ca să exemplificăm, în Standardele APA din anii 1954 şi 1966, fidelitatea însemna *consistenţa* sau *stabilitatea* ("Fidelitatea se referă la precizia – consistenţa şi stabilitatea – măsurării realizate de test"). În 1974, fidelitatea era definită ca fiind "gradul în care rezultatele testării pot fi atribuite surselor sistematice de varianţă". În Standardele de după 1985, apare o nouă schimbare în definirea fidelităţii, aceasta apărând ca fiind "gradul în care scorurile testului sunt consistente sau repetabile, adică gradul în care ele sunt afectate de erorile de măsură". Aşadar, aşa după cum precizează şi prof. H. Pitariu, termenii de consistenţă şi stabilitate continuă să fie utilizaţi, dar nu în sensul de consistenţă internă şi stabilitate temporară a unui anumit comportament, ci, mai degrabă, ca reproductibilitate a rezultatelor cercetării ştiinţifice.

Conceptul de fidelitate este utilizat pentru a acoperi mai multe aspecte ale consistenţei scorului (Anastasi, A., 1996). În primul rând, fidelitatea indică măsura în care diferenţele dintre subiecţii testaţi, evidenţiate prin intermediul scorurilor obţinute, se datorează diferenţelor "reale" (adevărate) privind însuşirea psihică măsurată şi măsura în care ele sunt atribuibile erorilor şansei. Să explicăm puţin mai mult. În urma aplicării testului, se obţine un anumit scor, numit scor brut (raw score), care poate fi exprimat cifric. De exemplu, la un test de inteligenţă se obţine scorul 120. În mod ideal, el ar trebui să reflecte exact realitatea, însă niciodatră nu se întâmplă aşa, deoarece, pe parcursul activităţii de testare pot interveni diverşi factori care scapă de sub controlul celui ce aplică testul. De aceea, niciodată scorul brut nu coincide cu scorul "real" al subiectului. Nu putem vorbi de utilizarea ecuaţiei X = SR, X însemnând scorul brut iar SR, scorul real. În realitate întâlnim situaţia evidenţiată de următoarea ecuaţie:

$$X = SR + E \quad \text{în care E = eroare}$$

Plecând de la această ecuaţie, putem spune: cu cât eroarea este mai mică, cu atât este de aşteptat mai mult ca scorul real să se apropie de scorul brut. Ce concluzie putem trage de aici ? Trebuie luate toate măsurile posibile pentru ca eroarea să fie cât mai mică. Orice condiţie, orice factor care este irelevant pentru scopul urmărit de un anumit test constituie o sursă pentru varianţa erorii. Astfel, când examinatorii încearcă să menţină uniforme condiţiile testării (respectând întocmai prevederile din cadrul manualului şi controlând

Copyright © DEPARTAMENT ID 2008

factorii ambianței în care se face testarea), ei reduc varianța erorii, ceea ce înseamnă că fac mult mai sigure scorurile obținute.

Apare, însă, o întrebare: are examinatorul posibilitatea să controleze toți factorii care pot interveni în timpul acțiunii de testare, căpătând statutul de "factor irelevant"? Răspunsul nu poate fi decât pozitiv, în practică existând cel puțin două situații: a) când controlul asupra factorilor irelevanți tinde către maximum, ceea ce înseamnă că scorul real se apropie până la identitate de scorul brut; b) când controlul asupra factorilor irelevanți tinde către minimum, ceea ce înseamnă că scorul real se îndepărtează foarte mult de scorul brut, eroarea în acest caz crescând foarte mult.

Dar care sunt aceşti factori irelevanți care pot să intervină în acțiunea de testare ? Unii autori au încercat să-i inventarieze. Astfel, de exemplu, făcând o sinteză a diferitelor lucrări de specialitate, H. Pitariu ne oferă următorul tablou al surselor posibile de variabilitate care pot afecta scorurile unui test.

I. Particularități durabile şi generale ale persoanei (cu influență asupra rezolvării mai multor teste:

A. Nivelul unor deprinderi care pot influența modul de a opera cu o serie de teste (de exemplu, un rezolvitor de cuvinte încrucişate va fi avantajat în rezolvarea unor teste).

B. Deprinderile şi tehnicile generale solicitate în testare (de exemplu, "test de excepție pentru profesionişti" sau "test pentru necunoscuți").

C. Abilitatea generală de a înțelege instrucțiunile de lucru cu testele (de exemplu, persoanele cu deficiențe intelectuale vor înțelege mai greu instrucțiunile de operare cu testul, la fel cele neobişnuite cu examinări prin teste).

II. Particularități durabile dar specifice ale persoanei (cu influență numai asupra unui test anume):

A. Specifice față de testul luat ca întreg

1. Nivelul individual de dezvoltare a abilității măsurate de acest test şi care nu influențează rezolvarea altor teste (de exemplu, o persoană dominantă va obține scoruri mari la scala de "Dominanță" din cadrul testului CPI, dar nu şi la scala de "Responsabilitate".

2. Cunoştințe şi deprinderi specifice formei particulare a testului. De exemplu, un programator va fi mai performant pe un test care solicită rezolvarea de scheme logice lacunare; sau, un ceasornicar va avea rezultate mai bune la un test de dexteritate digitală care solicită operarea cu obiecte mici);

3. Pattern-uri de răspuns stabile (de exemplu, la un test cu răspunsuri "Adevărat/Fals", tendința de a alege opțiunea "Adevărat" în toate situațiile de indecizie).

B. Specifice anumitor itemi ai testului:

1. "Şansa" unei persoane în a cunoaşte răspunsul la un anumit fapt particular cerut de un item (de exemplu, la un test de cunoştințe tehnice, un item poate solicita o informație pe care o cunosc doar subiecții care au citit o anumită lucrare).

2. Tipuri de itemi cu care diferiți subiecți sunt familiarizați inegal (de exemplu, existența unor itemi cu răspuns la alegere îi poate încurca pe cei nefamiliarizați cu stilul respectiv de lucru.

III. Particularități temporare, dar generale ale persoanei (este vorba de existența unor factori incidentali care pot afecta performanțele în momentul în care este administrat testul în cauză: a) sănătatea; b) oboseala; c) motivația; d) tensiunea emoțională; e) nivelul prea

Copyright © DEPARTAMENT ID 2008

elevat al testului; f) înţelegerea mecanismelor testării în cauză; g) condiţiile de administrare a testului, ca iluminat, căldură, umiditate, ventilaţie etc.

IV. Particularităţi temporare şi specifice ale persoanei (cu influenţă numai asupra acestui test):

A. Specifice faţă de testul luat ca întreg:

1. Înţelegerea sarcinii specifice cerute de testul în cauză (de exemplu, la un test administrat de calculator, un subiect care nu a reţinut modul de utilizare a tastaturii pentru formularea răspunsului poate obţine un scor care nu îl caracterizează).

2. Mici trucuri sau tehnici neortodoxe de rezolvare a testului (de exemplu, la un test de cunoştinţe alcătuit din itemi cu răspunsuri la alegere, care nu penalizează răspunsurile greşite, marcarea tuturor răspunsurilor propuse la fiecare întrebare conduce la obţinerea unui scor înalt).

3. Nivelul de antrenare în rezolvarea unor sarcini specifice cerute de testul în cauză (intervine mai ales în operarea cu teste psihomotoare; de exemplu, la un test de viteză de reacţie, administrat pe calculator, o persoană familiarizată cu tastatura calculatorului respective este avantajată).

4."Setul" sau dispoziţia momentană pentru rezolvarea testului (de exemplu, o persoană foarte ocupată în ziua testării poate să i se pară prea lung testul CPI, răspunzând, în consecinţă, la întâmplare).

B. Specifice faţă de unii itemi ai testului:

1. Fluctuaţii şi ideosincrazii ale memoriei umane (de exemplu, la un test de cunoştinţe de geografie, unui elev căruia nu-i place să opereze cu numere, nu-şi va aminti înălţimile unor munţi).

2. Fluctuaţii neprevăzute ale atenţiei sau percepţiei, suprapuse peste nivelul general de performanţă caracteristic unei persoane (de exemplu, la un test administrat de calculator, subiectul poate, din neatenţie, să apese pe o tastă greşită şi să nu observe acest lucru).

V. Factori sistematici sau de "noroc" care afectează administrarea testului sau evaluarea performanţelor la test:

A. Condiţiile de testare: utilizarea unui interval de timp fixat pentru testare, prezenţa unor factori de distragere a atenţiei, claritatea instrucţiei etc.

B. Efectul interacţiunii personalităţii, sexului sau rasei examinatorului cu persoana examinată, aceasta stimulând sau inhibând performanţa.

C. Inconsecvenţa sau părtinirea în evaluarea performanţelor

VI. Variaţii neexplicate

A. Norocul în găsirea răspunsului corect, când subiectul îl ghiceşte.

B. Distragerea momentană a atenţiei

Teme de reflecţie/autoevaluare:
1.Ce este fidelitatea şi ce este validitatea unui test ?
2. Ce înseamnă varianţa erorii şi cui se datorează ea ?

1. 3. Alte caracteristici psihometrice ale testelor

Aşa după cum am menţionat deja, în literatura anglo-saxonă singurele caracteristici psihometrice menţionate sunt fidelitatea şi validitatea. Unii autori, inclusiv autori români (de exemplu, U. Schiopu, 1974) vorbesc şi de o altă caracteristică psihometrică, respectiv, *fineţea discriminativă*. Aceasta s-ar referi, pe de o parte, la capacitatea testului de a

Copyright © DEPARTAMENT ID 2008

scoate în evidenţă diferenţele interindividuale şi, pe de altă parte, la capacitatea testului de a măsura exclusiv însuşirea psihică pentru care a fost construit. Alţi autori vorbesc de existenţa unui număr mare de asemenea caracteristici psihometrice. De exemplu, E. Claparède ne oferă o adevărată listă care cuprinde (desigur, el s-a referit, mai ales, la calităţile ideale pentru unele teste de aptitudini): 1. să fie interesante; 2. să fie obiective în aplicaţie; 3. să aibă un instructaj clar; 4. să manifeste obiectivitate în apreciere (notare); 5. să excludă hazardul; 6. să fie gradabile; 7. să fie etalonate; 8. să prezinte o bună dispersie; 9. să fie univoce (adică să măsoare un singur aspect odată); 10. să fie reprezentative; 11. să fie inedite; 12. să nu facă apel la cunoştinţe şcolare; 13. să fie constante; 14. să se aplice la toate vârstele; 15. să permită să se alcătuiască forme echivalente; 16. să nu ceară pe cât posibil aparatură; 17. să nu măsoare decât o singură variabilă (cantitate, rapiditate sau calitate).

Temă de reflecţie/ autoevaluare:
Citiţi cu atenţie cele 17 calităţi ale testelor şi stabiliţi dacă pot fi asimilate la cele două caracteristici psihometrice, respectiv, fidelitatea şi validitatea.

Întrebări de autoevaluare:

1. Ce sunt caracteristicile psihometrice ale testelor ?
2. Ce obligaţii au constructorii de teste în raport cu ele ?
3. Ce este fidelitatea testului ?
4. Ce este validitatea testului ?
5. Care sunt elementele de asemănare şi de deosebire dintre fidelitate şi validitate ?
6. Ce este scorul brut şi ce este scorul real al subiectului ?
7. Ce este varianţa erorii în măsurare?
8. Ce pericole prezintă varianţa erorii ?
9. Ce trebuie făcut pentru reducerea varianţei erorii ?
10. Enumeraţi cât mai mulţi factori irelevanţi ce pot interveni în acţiunea de testare.
11. Ce trebuie făcut pentru ca să reducem cât mai mult din acţiunea acestor factori irelevanţi ?
12. Ce putem spune despre controlul pe care trebuie să-l asigure cel ce aplică testul asupra factorilor irelevanţi ?

BIBLIOGRAFIE MINIMALĂ

ALBU, MONICA, PITARIU, HORIA (1993), *Proiectarea testelor de cunoştinţe şi examenul asistat de calculator*. Cluj-Napoca: Casa cărţii de ştiinţă.

ANASTASI, ANNE (1996), *Psychological Testing* (7th edition). New-York: Mcmillan.

CLINCIU, ION, AUREL (2005), *Psihodiagnostic*. Braşov: Editura Universităţii "Transilvania".

STAN, AUREL (2002), *Testul psihologic. Evoluţie, construcţie, aplicaţie*. Iaşi: Polirom.

HORGHIDAN, VALENTINA (1998), *Metode de psihodiagnostic*. Bucureşti: Editura didactică şi pedagogică.

Copyright © DEPARTAMENT ID 2008

Unitatea de învățare 2

Fidelitatea testului

Cuprins:

2.1. Proceduri de calculare a fidelității testelor

2.2. Eroarea standard a măsurării :

Obiective:
La sfârşitul acestei unități de învățare, studenții vor fi capabili să:
- cunoască tipurile de fidelitate
- cunoască procedurile de calcul a fidelității
- interpreteze exact ce înseamnă eroarea standard a măsurării
- dea exemple din practică

2.1. Proceduri de calculare a fidelității testelor

Mai întâi, să le enumerăm: a) fidelitatea test-retest; b) fidelitatea formă alternantă; c) fidelitatea half-split; d) fidelitatea Kuder-Richardson; e) fidelitatea coeficient alpha; f) fidelitatea privind personalitatea examinatorului. Dar de ce sunt aşa de multe ? În primul rand, din cauză că, aşa cum am văzut la clasificarea testelor, acestea sunt foarte diferite (de exemplu, un test de inteligență diferă total de un chestionar de personalitate sau de un test proiectiv). Desigur, nu este exclusă situația în care pot fi utilizate mai multe proceduri pentru acelaşi test şi atunci se pune problema calculării fidelității totale.

a) *fidelitatea test-retest*. Din punct de vedere strict tehnic, această modalitate presupune aplicarea unui test, în momentul t_1, pe un grup de subiecți şi apoi, reaplicarea lui, pe acelaşi grup de subiecți, în momentul t_2. În continuare se calculează valoarea lui r_{tt}, adică a coeficientului de fidelitate, care arată la ce nivel corelează cele două serii de rezultate. Distanța dintre cele două momente poate fi mai mică, de ordinul zilelor sau săptămânilor, sau poate fi mai mare, de ordinul lunilor sau chiar anilor. În urma multor ani de cercetări s-a constatat că intervalele de timp mai scurte asigură o valoare mai mare a fidelității, în timp ce intervalele mari nu mai asigură aproape niciun fel de corespondență între cele două serii de rezultate. Oricum, aplicarea testului în cele două momente înseamnă şi aplicarea lui în condiții oarecum diferite. Pot interveni anumiți factori irelevanți (vă mai amintiți de ei, desigur !!) care scapă de sub controlul experimentatorului, făcând să crească valoarea erorii în măsurare. Aceşti factori pot ține de ambianța în care se face măsurarea (pot exista diferențe de temperatură, surse de distragere a atenției, diferențe privind spațiul etc.) sau de persoana subiecților (oboseală, experiențe

Copyright © DEPARTAMENT ID 2008

personale diferite, stări de dispoziție diferite etc). Deci, varianța erorii corespunde fluctuațiilor întâmplătoare ale performanței de la o testare la alta.

Când intervalul de timp este scurt sau foarte scurt (de ordinul zilelor, de exemplu) se poate întâmpla ca subiecții să fie puternic influențați de ceea ce înseamnă recunoaşterea "principiului de construcție a testului". Să luăm, mai întâi, câteva cazuri simple, pentru a înțelege mai bine:

Exemplul 1. Itemul testului apare în forma următoare:

A		R		P
-	duc	-	un	-
L		D		?

Sarcina subiectului este de a spune ce literă trebuie pusă în locul semnului de întrebare. Voi ce părere aveți ?

Exemplul 2. Itemul testului apare în forma următoare:

COCOLOŞ	masă
BALAMA	cobai
VIŞINE	praline
PEPENE	oaie

Sarcina subiectului este de a stabili cu care dintre cuvintele scrise cu litere mici se asociază cuvintele scrise cu litere mari.

Exemplul 3. Itemul testului apare în forma următoare:

LONDRA este oraşul în care se găssesc 6 teatre, 4 stadioane şi 24 de cinematografe

MOSCOVA este oraşul în care se găsesc 3 stadioane, 21 de cinematografe şi 7 teatre

PARIS, capitala Franței, este oraşul în care se găsesc 25 de cinematografe, 5 teatre şi 5 stadioane

Sarcina subiectului: câte teatre are oraşul ROMA ?

Vom comenta acest ultim exemplu, lăsându-vă bucuria de a descoperi singuri care este răspunsul la primele două cazuri. Principiul de construcție al itemului constă în asocierea dintre numărul de litere ce-l regăsim în denumirea oraşului şi numărul de teatre. LONDRA are 6 teatre, pentru că are şase litere, MOSCOVA are 7 teatre, pentru că are 7 litere şi Paris are 5 teatre, pentru că are 5 litere. Deci, ROMA are 4 teatre, deoarece are 4 litere. Celelalte elemente din structura itemului (numărul de stadioane şi de cinematografe) sunt introduce pentru a face sarcina puțin mai dificilă pentru subiect.

Acum ne dăm seama cât de mult ar fi ajutat subiectul, cunoscând principiul de construcție al testului, dacă el ar trebui să fie reexaminat cu acelaşi instrument la un interval de timp scurt.

Copyright © DEPARTAMENT ID 2008

Pe de altă parte, în cazul în care intervalul de timp este foarte mare, este greu de ignorat impactul pe care îl are asupra subiectului experiența lui personală (noi achiziții în plan informațional-cognitiv, în plan comportamental etc.).

În practica psihodiagnostică s-au conturat deja un fel de reguli sau recomandări şi anume: - în general, pentru orice tip de persoană intervalul dintre cele două momente de testare (test-retest) nu trebuie să depăşească, decât foarte rar, 6 luni; - pentru copiii mici şi foarte mici, datorită schimbărilor accentuate ca urmare a dezvoltării progresive, intervalul dintre cele două momente ale testării poate şi trebuie să fie mai scurt decât în cazul adulților.

Deşi această procedură de calculare a fidelității este foarte mult folosită, există, totuşi, recomandarea ca să se apeleze la ea, mai ales în cazul testelor care nu pot fi afectate de repetiție. În asemenea categorie ar intra unele teste de discriminare senzorială şi unele teste psihomotorii (Anastasi, A., 1996).

b) *fidelitatea formă alternantă*. În mod practic se procedează în felul următor: se aplică pe un grup de subiecți un test şi, apoi, după un anumit interval de timp, care poate fi mai scurt sau mai lung, se aplică, pe acelaşi grup, un alt test dar care este obligatoriu *echivalent* cu primul. În final, se calculează valoarea coeficientului de fidelitate, respectiv, r_{tt}. În cazul în care forma alternantă a testului se aplică imediat după ce s-a aplicat prima variantă a lui, vorbim de fidelitate formă alternantă imediată iar, în cazul în care intervalul de timp este mai mare, vorbim de fidelitate formă alternantă întârziată. În primul caz, sursa variației erorii o constituie eşantionarea conținutului şi, în al doilea caz, sursa variației erorii o constituie eşantionarea conținutului dar şi eşantionarea timpului.

În cadrul manualului testului trebuie să fie incluse, în mod obligatoriu, informații privind lungimea intervalului de timp dintre cele două momente de administrare şi, în cazul în care acest interval este mai mare, informații privind tipurile de experiențe cu care s-au confruntat subiecții.

Dar cea mai importantă problemă pe care o ridică folosirea acestei proceduri se referă la modul în care se asigură ca cele două variante ale testului să fie *identice*, deci se pune problema asemănărilor şi deosebirilor. În practică este destul de dificil de realizat acest lucru, dar nu este imposibil, dovadă că există asemenea cazuri. Principala deosebire constă în faptul că ele nu seamănă atunci când sunt puse în fața subiecților. Sunt alte figuri, alte exerciții, alte întrebări etc. Asemănările însă sunt mai multe şi ele se referă la: - obiectivul urmărit în acțiunea de măsurare este acelaşi; - privitor la conținut, există un număr egal de itemi; - acelaşi tip de instructaj, acelaşi format al testului; - egalitate în ceea ce priveşte nivelul de dificultate al itemilor

c) *fidelitatea half-split*, sau fidelitatea bazată pe divizarea testului în două jumătăți. Din punct de vedere tehnic, testul este împărțit în două, cele două jumătăți devenind un fel de teste independente, se aplică pe un grup de subiecți şi apoi se calculează valoarea coeficientului de fidelitate (r_{tt}). Dar cum realizăm cele două jumătăți? Vom oferi două modalități: 1) împărțirea pur şi simplu a testului în două părți, fiecare conținând un număr egal de itemi. De exemplu, dacă testul întreg are 100 de itemi, cele două jumătăți vor avea câte 50 de itemi. Se aplică testul pe grupul de subiecți şi, apoi, se calcuează coeficientul de corelație (respectiv, coeficientul de fidelitate) dintre cele două serii de rezultate. Sunt însă unele teste pentru care această modalitate nu poate fi aplicată. Astfel, dacă testul este construit pe principiul introducerii itemilor de la uşor la greu, atunci prima jumătate va fi foarte uşoară şi cealaltă jumătate va fi foarte grea. Sau, dacă testul este foarte lung, are un

Copyright © DEPARTAMENT ID 2008

număr mare de itemi, la aplicarea celei de-a doua jumătăţi va interveni în mod sigur oboseala;

2) o procedură mai adecvată presupune împărţirea testului în două jumătăţi, plecând de la numărul de ordine al itemilor. În prima jumătate sunt reţinuţi itemii cu soţ (2,4,6,8 ş.a.m.d.) şi în a doua jumătate, itemii cu soţ (1,3,5,7 ş.a.m.d.). Astfel, sunt mai multe şanse ca cele două jumătăţi să fie echivalente. Desigur, o semenea procedură nu se poate aplica în cazul în care mai mulţi itemi se referă la o aceeaşi problemă, cum ar fi cazul unei situaţii problematice.

Dar indiferent de ce procedură folosim, valoarea coeficientului de corelaţie vizează fidelitatea doar a unei jumătăţi de test, or, pe noi ne interesează fidelitatea pentru întregul test. O primă tentaţie ar fi aceea de a dubla, pur şi simplu, valoarea coeficientului de fidelitate. Dacă am proceda aşa, am ajunge la rezultate stranii. Gândiţi-vă, de exemplu, dacă valoarea lui r_{tt} pentru jumătate de test ar fi 0,65, dublând am ajunge la o valoare egală cu 1,30. Poate avea coeficientul de corelaţie o asemenea valoare ? V-am rugat să vă întoarceţi la ce aţi studiat la Statistică….

În realitate se aplică formula lui Spearman-Brown (Anastasi, A., 1988, 1996):

$$r_{tt} = \frac{2r_{hh}}{1 + r_{hh}}$$

în care: r_{tt} = coeficientul de fidelitate
r_{hh} = corelaţia privind jumătatea de test

Să luăm un exemplu: dacă $r_{hh} = 0,50$, aplicăm formula şi vom vedea care este valoarea coeficientului de fidelitate pentru întregul test:

$$r_{tt} = \frac{2 \times (0,50)}{1 + (0,50)} = \frac{1}{1,5} = 0,66$$

Acum voi trebuie să faceţi câteva exerciţii şi să vedeţi ce se întâmplă cu valoarea coeficientului de fidelitate, când creşte sau scade valoarea coeficientului de fidelitate pentru jumătate de test: 1. $r_{hh} = 0,30$; $r_{hh} = 0,80$; $r_{hh} = 0,99$.

Efectul pe care creşterea sau scăderea numărului de itemi ai testului îl are asupra valorii coeficientului de fidelitate poate fi calculat cu ajutorul altei formule Spearman-Brown:

$$r_{nn} = \frac{nr_{tt}}{1 + (n-1) r_{tt}}$$

în care: r_{nn} = coeficientul estimat
r_{tt} = coeficientul de corelaţie obţinut
n = numărul ce exprimă de câte ori creşte sau descreşte dimensiunea testului

Să luăm un exemplu: dacă dorim ca numărul itemilor unui test, al cărui coeficient de fidelitate are valoarea 0,50, să crească de la 25 la 100, deci să crească de 4 ori, aplicăm formula:

$$r_{nn} = \frac{4 \times (0,50)}{1 + (3 \times 0,50)} = \frac{2}{2,5} = 0,80$$

Copyright © DEPARTAMENT ID 2008

Deci, putem uşor constata că, în cazul în care mărim dimensiunile unui test, valoarea coeficientului de fidelitate creşte. Şi, de aici, o concluzie foarte importantă pentru practică: cu cât numărul itemilor unui test este mai mare, cu atât nivelul de credibilitate acordat informațiilor obținute cu ajutorul lui va creşte. Am putea folosi o comparație: dacă am asemăna itemii cu ochii, cu cât sunt mai mulți ochi cu atât va fi "văzută" mai bine o anumită realitate.

Pentru a vedea dacă se confirmă ceea ce spunem, încercați să faceți şi voi unele exerciții şi anume: care este valoarea lui r_{nn} dacă: 1. dorim să micşorăm dimensiunile unui test (r_{tt} = 0,50) de 2 ori; 2. dorim să micşorăm dimensiunile unui test (r_{tt} = 0,50) de 5 ori şi de 10 ori.

Se poate pune problema şi aşa: cunoaştem valoarea actuală a coeficientului de fidelitate a unui test (care, normal, are un număr determinat de itemi) şi dorim ca această valoare să fie îmbunătățită, adică să atingă o anumită valoare; avem nevoie să ştim care va trebui să fie numărul itemilor pentru testul modificat. În acest caz, aplicăm formula profeției Spearman-Brown:

$$N = \frac{rtt_d\,(1 - rtt_o)}{rtt_o\,(1 - rtt_d)}$$

în care: rtt_o = coef. de fidelitate observat
rtt_d = coef. de fidelitate dorit
N = de câte ori va creşte numărul itemilor

Să luăm un exemplu: un test are 20 de itemi şi rtto = 0,87; dacă dorim ca rttd să aibă valoarea 0,95, până la ce nivel trebuie să crească numărul itemilor ? Aplicăm formula:

$$N = \frac{0,95\,(1 - 0,87)}{0,87\,(1 - 0,95)} = 2,82$$ Apoi, 20 x 2,82 = 56,4, rotunjit, 56

Deci, pentru ca valoarea coeficientului de fidelitate să atingă nivelul de 0,95, numărul itemilor testului trebuie să atingă nivelul de 56.

Dar de ce este important ca valoarea fidelității testului să fie cât mai mare ? Desigur, cazul ideal este ca valoarea lui rtt să fie egală cu +1,00. În realitate nu găsim asemenea cazuri, valoarea lui rtt fiind mai aproape sau mai departe de această valoare ideală. Testele existente în practică au valori diferite ale fidelității, ridicându-se problema selecției lor în funcție de ce obiective urmărim să atingem folosind rezultatele obținute. În acest sens, unii autori fac următoarele recomandări (Kaplan, R., Saccuzzo, D., 1993): a) pentru cele mai multe din scopurile de cercetare, valoarea acceptată a fidelității testelor folosite este de minimum 0,70; b) când testul este folosit pentru a lua o decizie importantă în legătură cu o persoană, rtt trebuie să aibă o valoare mai mare de 0,95; c) în alte scopuri de testare, valoarea acceptată a fidelității este de minimum 0,90.

d) *fidelitatea Kuder-Richardson*. Această procedură este aplicabilă mai ales pentru testele ale căror itemi sunt prevăzuți cu un sistem de răspuns de tip "tot sau nimic". Este vorba de variantele disjunctive, de genul: "Da – Nu", "Adevărat-Fals" etc. Ceea ce este specific pentru această procedură este faptul că ea este bazată pe performanța la fiecare item. Formula care se aplică este următoarea:

$$r_{tt} = \frac{n}{\quad} \times \frac{SD_t^2 - \Sigma pq}{\quad}$$, în care.

Copyright © DEPARTAMENT ID 2008

$$\frac{(n-1)}{} \qquad \frac{}{SD_t^2}$$

r_{tt} = coeficientul de fidelitate al întregului test
n = numărul de itemi ai testului
SD_t = deviația standard a scorurilor totale ale testului
Σpq = suma produselor persoanelor care reuşesc şi a celor care nu reuşesc la fiecare item

Din punct de vedere practic, se aplică un test pe un grup de subiecți, apoi produsul *pq* se calculează pentru fiecare item şi apoi toate produsele sunt adunate, rezultând Σpq. Se calculează SD_t, se ridică la pătrat şi se completează datele cerute de formulă.

Să luăm şi noi un exemplu practic (Friedenberg, L., 1995): unui grup de studenți i s-a aplicat un test ce cuprinde 6 întrebări. Răspunsul corect a fost marcat cu litera "D" iar răspunsul incorect, cu litera "N". În rubrica "scor total" regăsim numai numărul răspunsurilor corecte.

Întrebare	1	2	3	4	5	6	Scor total
A	D	D	D	D	N	D	5
B	D	N	N	D	N	D	3
C	D	N	N	D	N	N	2
D	N	D	N	N	N	N	1
E	N	D	N	N	D	D	3

p =	0,6	0,6	0,2	0,6	0,2	0,6	
q =	0,4	0,4	0,8	0,4	0,8	0,4	
pq =	0,24	0,24	0,16	0,24	0,16	0,24	$\Sigma (p)(q) = 1,28$

($\bar{X} = 2,8$; $SD^2 = 1,76$; $SD = 1,3266$; $n = 6$; $n = 5$) Aplicând formula:

$$KR\text{-}20 = \frac{6}{5}\left(\frac{(1,76) - 1,28}{(1,76)}\right) = (1,2)\ \frac{0,48}{1,76} = (1,2)(0,2727) = 0,33$$

e) *coeficientul alpha*. În cazul altor teste, cum ar fi testele de personalitate, răspunsul nu mai poate fi disjunctiv, de tipul "Da" – "Nu", ci subiectul trebuie să aleagă dintre mai multe răspunsuri. De exemplu, la un item de genul "Obişnuiți să consumați băuturi alcoolice ?", subiectul trebuie să aleagă din mai multe răspunsuri la alegere care sunt prezentate pe o scală cu 3 trepte, cu 5 trepte. Astfel, la întrebarea de mai sus, scala cu 3 trepte ar fi aşa : niciodată – uneori – mereu, iar scala cu 5 trepte ar arăta aşa:

niciodată - foarte rar - uneori - foarte des - întotdeauna

Fiecare variantă primeşte un anumit punctaj; legat de exemplul de mai sus, dacă urmărim să vedem în ce măsură subiectul este dependent de alcool, varianta de răspuns "întotdeauna" va primi punctaj maxim, respectiv, 5 puncte şi, apoi, în ordine descrescătoare: "foarte des" – 4 puncte; "uneori – 3 puncte; "foarte rar" – 2 puncte; "niciodată" – 1 punct.

Copyright © DEPARTAMENT ID 2008

Pentru această categorie de instrumente psihodiagnostice formula generalizată şi propusă de Cronbach (1951), Kaiser şi Michael (1975), Novik şi Lewis (1967) ajută la calcularea *coeficientului alpha*, care este, după cum am mai spus, coeficient de fidelitate. Formula care se poate folosi este următoarea:

$$r_{xx} \text{ (sau } \alpha) = \frac{n}{n-1}\left(1 - \frac{\Sigma(SDi)^2}{SD_x^2}\right) \text{ în care:}$$

SD_x^2 = suma varianţelor la scorurile întregului test
$(SDi)^2$ = suma varianţelor la scorurile fiecărui item al testului
n = numărul de itemi ai testului

Item	1	2	3	4	5	Scor total
Sub. 1	3	4	4	3	5	19
Sub. 2	4	3	4	3	3	17
Sub. 3	2	3	3	2	3	13
Sub. 4	4	4	5	3	4	20
Sub. 5	3	2	4	3	3	15
Sub. 6	3	2	3	2	3	13

Deci, din punct de vedere practic, această procedură presupune găsirea varianţei la scorurile subiecţilor pentru fiecare item, apoi se face suma, ajungându-se la $\Sigma(SDi)^2$. Să luăm şi noi un exemplu practic (Friedenberg, L., 1995): unui grup de 6 elevi i s-a aplicat un test, tip eseu, alcătuit din 5 întrebări. La fiecare dintre întrebări scorul maxim era de 5 puncte.

$SD_i^2 =$ .4722 .6667 .4722 .2222 .5833

(X = 16.1667; SD^2 = 7.4722; SD = 2.7335; n = 5; N = 6)

Aplicăm formula de mai sus:

$$A = \frac{5}{4}\left(1 - \frac{.4722 + .6667 + 4722 + .2222 + .5833}{7.4722}\right) =$$

$$= (1.25)\left(1 - \frac{2.4166}{7.4722}\right) = 1.25)\,(1 - .3234) = .84575 = .85$$

f) *fidelitatea legată de persoana celui care aplică testul*. Din punct de vedere practic, după ce se aplică un test, mai mulţi examinatori sunt puşi să coteze rezultatele. Apoi se aplică procedurile uzuale de calculare a coeficientului de corelaţie şi ceea ce rezultă este

Copyright © DEPARTAMENT ID 2008

valoarea fidelității examinatorului. Desigur, este vorba, mai ales, de acele teste care intră în categoria testelor subiective (vă mai amintiți de clasificarea testelor în teste obiective şi teste subiective ?) şi, mai exact este vorba de testele proiective, la care diferențele dintre modul de cotare a aceloraşi rezultate de către mai mulți examinatori pot fi destul de mari. Iar dacă, pentru un anumit test, se calculează valoarea acestui tip de coeficient de fidelitate, acest gen de informație trebuie să fie inclus în cadrul manualului testului.

După ce am trecut în revistă mai multe forme ale fidelității, să vedem, sintetic, care sunt sursele specifice ale varianței erorii pentru ele (Anastasi, A.,1988, 1996).

Nr. crt.	Tipul de coefficient de fidelitate	Surse pentru varianța erorii
1	Fidelitate test-retest	- eşantionarea timpului
2	Fidelitate formă alternantă (imediată)	- eşantionarea conținutului
3	Fidelitate formă alternantă (la un interval de timp	- eşantionarea timpului şi a conținutului
4	Fidelitate half-split	- eşantionarea conținutului
5	Kuder-Richardson şi Coeficientul Alpha	- eşantionarea conținutului şi eterogenitatea conținutului
6	Fidelitate legată de examinatori	- diferențele dintre examinatori

Aşa după cum am mai precizat, valoarea coeficientului de fidelitate arată care este procentul din varianța scorurilor la test ce depinde de varianța reală legată de trăsătura supusă măsurării. De exemplu, dacă r_{tt} are valoarea 0,75, asta înseamnă că 75% reprezintă varianța reală şi 25%, varianța erorii.

Tema de reflecție/autoevaluare

Vi s-au recomandat să faceți câteva exerciții. Ce ați putut constata în urma efectuării lor ?

2.2. Eroarea standard a măsurării.

Este o altă formă de măsurare a fidelității, dar care este folosită pentru a interpreta scorurile individuale. De aceea, mai este denumită şi eroarea standard a scorului. Ca formulă de calcul este folosită următoarea:

$SEM = SD_t \sqrt{1 - r_{tt}}$ în care: SD_t = deviația standard a scorurilor testului

r_{tt} = coeficientul de fidelitate

De exemplu, dacă vom cunoaşte valoarea deviației standard, precum şi a coeficientului de fidelitate, putem calcula uşor valoarea lui SEM. Astfel, dacă SDt = 10 şi rtt = 0,80, valoarea lui SEM va fi 5. De regulă, constructorul de test se ocupă şi de această problemă iar valoarea obținută a lui SEM este inclusă în cadrul manualului. Pe noi ne interesează mai mult să vedem ce reprezintă această valoare şi la ce ne foloseşte. După cum am mai afirmat, scorul obținut de un subiect la un test (scor observat), cuprinde scorul real (raw score) şi eroarea. Un test este cu atât mai valoros cu cât eroarea este mai mică. Datorită erorii, în cazul în care reaplicăm testul pe un subiect, scorul va varia, va fi mai mare sau mai mic şi cât de mult poate varia ne spune valoarea lui SEM.

Copyright © DEPARTAMENT ID 2008

Să luăm un exemplu şi mai concret din literatura de specialitate (Friedenberg, L., 1995), cu ajutorul căruia să ilustrăm şi mai bine conceptual de SEM. Un subiect este testat cu un test de inteligență de un număr mare de ori, prima dată el obținând un QI = 120. Testarea se va realiza, desigur, la anumite intervale de timp, ceea ce înseamnă că vor interveni, în mod sigur, o serie de factori irelevanți (vă mai amintiți de ei ?), cum ar fi: condiții de testare diferite, starea subiectului diferită, poate fi mai oboist, poate fi mai necăjit, fluctuații ale capacității de concentrare a atenției etc. Deci, scorul poate fi mai mare sau mai mic şi, în consecință, dacă am avea 100 de testări, ne vom aştepta ca jumătate dintre scoruri să fie mai mari şi jumătate să fie mai mici. Apelând la curba lui Gauss (dacă nu ştiți despre ce este vorba, revedeți de urgență ce ați învățat la statsitică !!!), vom fi puşi în fața unei distribuții normale, având media = 120. SEM va reprezenta deviația standard a acestei distribuții aşteptate. În consecință: 34% din scorurile subiectului vor cădea între 120 şi + 1 SEM (însemnând media şi o deviație standard) şi alte 34% din scoruri vor cădea între 120 şi - 1 SEM (însemnând media şi - o deviație standard). Luate împreună, 68% din scorurile subiectului vor cădea în intervalul definit de 120 ± 1 SEM. Deci, noi putem fi siguri că, în caz de reaplicare a testului, 68% din scorurile subiectului vor cădea în acest interval, numit "interval de încredere de 68%"

Pentru a obține un interval de încredere de 95%, trebuie să determinăm numărul de deviații standard ce cuprind 95% din scorurile subiectului. Deci acest interval s-ar întinde între 120 şi ± 2 SEM. Deci, revenind la exemplul de mai sus, intervalul va fi între 120 + 10 şi 120 – 10, respectiv, între 130 şi 110. Mai departe, pentru a obține un interval de încredere de 99%, trebuie să determinăm numărul de deviații standard ce curpind 99% din scorurile subiectului. Şi acest interval s-ar întinde între 120 şi ± 3 SEM, respectiv, în cazul concret, între 105 şi 135.

Deci formula pentru construirea intervalului de confidență este următoarea:

X ± (scorul z) (SEM), în care:

X + (scorul z) (SEM) = limita superioară a intervalului

X - (scorul z) SEM = limita inferioară a intervalului

Teme de reflecție/autoevaluare:

1. Care este utilitatea practică a cunoaşterii valorii fidelității şi valorii erorii standard a măsurării (SEM) ?
2. Dați 3 exemple concrete de utilizare a erorii standard a măsurării.

Întrebări de autoevaluare:

1. Care sunt procedurile de calculare a fidelității testelor ?
2. Care este specificul fidelității test-retest ?
3. Ce se întâmplă când intervalul dintre cele două testări este prea mic sau este prea mare ?
4. Care este specificul fidelității formă alternantă ?
5. Ce reguli trebuie respectate pentru a realiza teste echivalente ?
6. Care este specificul fidelității half-split ? Cum se procedează pentru a afla valoarea fidelității pentru întregul test ?
7. Care este specificul fidelității Kuder-Richardosn şi Coeficientul Alpha ? Care este deosebirea dintre ele ?
8. Care este specificul fidelității legate de persoana examinatorului ?
9. Care sunt valorile fidelității recomandate pentru anumite scopuri ale testării ?

Copyright © DEPARTAMENT ID 2008

10. Care sunt sursele varianței erorii pentru dintre procedurile de calcul a fidelității ?
11. Ce reprezintă eroarea standard a măsurării (SEM) ?
12. La ce ne ajută cunoaşterea valorii lui SEM ?
13. Dați exemple privind utilizarea valorii lui SEM

BIBLIOGRAFIE MINIMALĂ

1. MINULESCU, M. (2003), *Teorie şi practică în psihodiagnoză*. Bucureşti: Editura Fundației *România de mâine*.
2. STAN, A. (2002), *Testul psihologic. Evoluție, construcție, aplicații*. Iaşi: Polirom.
3. CLINCIU, AUREL, ION (2005), *Psihodiagnostic*. Braşov: Editura Universității „Transilvania”.
4. KAPLAN, M. ROBERT (2005), *Psychological Testing. Principles, Applications, And Issues*. Wadsworth: Thompson.
5. FRIEDENBERG, L. (1995), *Psychological Testing. Desigh, Analysis, and Use*. Allyn & Bacon.

Copyright © DEPARTAMENT ID 2008

Unitatea de învățare 3
Validitatea testului

Cuprins
1. Validitatea de conținut
2. Validitatea de criteriu
3. Validitatea de construct

Obiective:
La sfârşitul acestei unități de învățare, studenții vor fi capabili să:

- cunoască noțiunea de validitate de conținut;
- cunoască noțiunea de validitate de criteriu;
- exemplifice cât mai multe categorii de criterii;
- cunoască noțiunea de validitate de construct;
- interpreteze valoarea coeficientului de validitate;
- cunoască procedurile de calculare a validității

1. Validitatea de conținut.

Se referă mai ales la categoria testelor de achiziție. Spunem că un test are validitate de conținut, dacă el măsoară ceea ce şi-a propus să măsoare şi dacă elementele sale de conținut (itemii) sunt expresia unui eşantion reprezentativ pentru un anumit univers de itemi sau univers de sarcini. Să ne explicăm mai mult. Să ne amintim de zilele cînd aveam de dat teză la o anumită disciplină. Pentru aceasta trebuia să citim (si să retinem !!!) 50 de pagini. Un elev s-a prezentat la teză cunoscând 45 de pagini iar altul doar 5 pagini, respectiv, cele pe care nu le citise primul. La teză s-au dat subiecte din cele 5 pagini. Primul elev nu a luat notă de trecere iar al doilea a luat o notă mare. Dacă această modalitate o asemănăm cu un test de cunoştințe, este clar că el nu are validitate de conținut. Subiectele au fost extrase la întâmplare iar norocul l-a favorizat pe al doilea elev. Nota obținută de cei doi elevi nu reflectă nivelul de achiziționare real al celor doi elevi. Dacă testul ar fi fost alcătuit ținându-se seama şi de cerințele validității de conținut, el ar fi trebuit să cuprindă itemi referitor la întreaga materie iar răspunsul dat de cei doi elevi ar fi reflectat corect nivelul de achiziționare a cunoştințelor.

Iată, aşadar, că este extrem de important ca, atunci când construim un test de cunoştințe, să-i asigurăm validitatea de conținut. Aceasta înseamnă mai multe aspecte. În primul rând, cei care fac o evaluare a programului educațional trebuie să fie experți în domeniu. Numai aceştia au competența de a analiza şi de a evalua elementele de conținut, formulând un anumit univers de itemi. Aceştia vor fi diferiți în funcție de gradul de dificultate şi, mai ales, în funcție de categoria de care aparțin. Unii se pot referi la cunoştințe factuale, alții se pot referi ca cunoaşterea unor principii, alții, la capacitatea de înțelegere şi de interpretare, alții, la capacitatea de evaluare etc. Deci, mai întotdeauna se pleacă de la obiectivele urmărite de către programul educațional. Acestea pot fi de natură informațională (elevul va cunoaşte, va şti....) şi de natură operațională (elevul va avea capacitatea de a face ceva....). În structura testului vom găsi itemi care vor evalua achizițiile realizate pe cele două direcții.

Copyright © DEPARTAMENT ID 2008

In concluzie, în baza informațiilor recoltate de către experți privitoare la conținutul programelor educaționale (cursuri, manuale, programe etc.) se conturează *specificațiile* testului, care trebuie să arate ariile de conținut sau subiectele ce trebuie să fie acoperite, obiectivele instrucționale (educaționale) sau procesele ce urmează a fi testate, precum şi importanța relativă a subiectelor şi proceselor individuale. Pe această bază se stabileşte numărul itemilor pentru fiecare tip de subiect, precum şi punctele acordate pentru răspuns corect (Anastasi, A., 1988, 1996).

Desigur, cei care construiesc noi teste de achiziție au obligația de a include, în cadrul manualului testului, informații privind procedurile utilizate, ariile de conținut, tipurile de achiziții (cunoştințe, deprinderi, capacități) acoperite de test, categoriile de itemi şi numărul de itemi pentru fiecare.

Referitor la testele de achiziție am tot vorbit despre validitatea de conținut. Asta nu înseamnă că şi alte tipuri de validitate nu pot fi folosite şi nu sunt relevante pentru aceste teste. Pe de altă parte, nu trebuie să se înțeleagă faptul că validitatea de conținut este aplicabilă numai pentru testele de achiziție. Astfel, în practica psihodiagnostică, acest tip de validitate este folosită în cazul testelor ocupaționale, care, la rândul lor, pot fi utilizate în activitatea de selecție profesională.

Şi ar mai fi o problemă legată de validitatea de conținut. În ce măsură avem garanția că experții (deci mai multe persoane) procedează identic în analiza şi evaluarea elementelor de conținut ? De aceea, unii autori consideră că este necesar să analizăm calitatea judecății lor prin raportare la anumite repere (Stan, A., 2002). Sintetic se pot rezuma trei repere: 1. Coerența internă a judecăților stabileşte persistența gradului de exigență a judecătorului de-a lungul efectuării actului apreciativ. Un judecător nu poate emite aprecieri foarte exigente doar pentru o parte din itemi, ci pentru ansamblul sarcinilor din testul supus analizei; 2. Varianța judecăților: între judecățile emise de diferiți experți nu trebuie să existe o prea mare diferență; 3. Concordanța judecăților: între judecățile emise de diferiți experți nu trebuie să existe o prea mare diferență; 3. Concordanța judecăților se realizează în cazul în care acestea au tendința de a ordona itemii în aceeaşi manieră, după gradul lor de congruență cu fațetele supuse măsurării.

A. Stan menționează că mai mulți autori au propus diferiți indicatori de măsurare a validității de conținut. Astfel, Lawshe a propus următoarea formulă pentru calcularea unui coeficient de validitate de conținut, CVR (inițialele, în limba engleză, de la *content validity ratio*):

$$CVR = \frac{N_e - \frac{N}{2}}{\frac{N}{2}}$$ formulă în care:

N_e = numărul evaluatorilor (experților) care consideră testul, respectiv itemul, ca fiind reprezentativ; N = numărul total de evaluatori (experți)

Temă de reflecție/autoevaluare:

1. În ce măsură rezultatele (performanțele) obținute la un test de cunoştințe sunt independente de influența variabilelor irelevante ?
2. Puteți să explicați de ce validitatea de conținut este nepotrivită pentru testele de personalitate şi testele de aptitudini ?

Copyright © DEPARTAMENT ID 2008

2. Validitatea de criteriu

Când am discutat despre funcţiile psihodiagnozei am arătat că una dintre cele mai importante este cea prognostică, adică în baza rezultatelor la test putem să facem predicţii privind conduita pe care o va adopta subiectul într-o activitate sau situaţie viitoare. Rezultatele obţinute de către subiec la test poartă numele de *predictori* iar rezultatele (performanţele) obţinute în cadrul activităţii (situaţiei) în raport cu care s-au făcut predicţiile poartă numele de *criteriu*. Deci, aici este vorba despre două mijloace de măsurare, respectiv, testul, în baza căruia facem predicţiile şi criteriul, respectiv, mijlocul de măsurare directă şi independentă a ceea ce a anticipat sau a prognosticat testul. Deci testul poate să anticipe corect conduita subiectului sau, dimpotrivă, nu poate face acest lucru. Puterea lui de anticipare sau de predicţie depinde de validitatea sa de criteriu, exprimată de coeficientul de validitate (r_{xy}). Cu cât valoarea acestuia este mai mare, cu atât devine mai sigură predicţia privind comportamentul viitor al subiectului. Desigur, un test poate fi validat printr-un singur criteriu sau prin mai multe criterii.

Legat de intervalul de timp dintre cele două momente de măsurare (test şi criteriu), acesta poate fi mai mare sau, dimpotrivă, mai mic, uneori atât de mic încât măsurarea criteriului se poate face în acelaşi timp cu măsurarea făcută cu ajutorul testului. În consecinţă, putem diferenţia două forme de validitate, respectiv, *validitate concurentă* şi *validitate predictivă*. În cazul validităţii concurente, deşi aplicăm un anumit test cu care măsurăm o anumită capacitate psihică (de exemplu, inteligenţa), rezultatele la criteriu le avem deja la îndemână, cum este cazul rezultatelor şcolare. În asemenea situaţii validitatea concurentă apare ca un substitut al validităţii predictive, ea vizând în special statutul existent al subiecţilor. Având la îndemână informaţiile privind rezultatele la test există pericolul influenţării celor care sunt implicaţi în evaluarea legată de criteriu. Acest fenomen este cunoscut sub numele de *contaminare a criteriului*. De exmplu, un cadru didactic, dacă ar cunoaşte rezultatele la teste ar manifesta tendinţa (este real acest pericol !!!) de a evalua la criteriu (cum este cazul unui test de cunoştinţe) în funcţie de aceste rezultate. De aici se desprinde concluzia clară că este bine ca cei ce evaluează la criteriu să nu aibă acest la rezultatele la teste.

Printre cele mai cunoscute criterii ce sunt utilizate în practica psihodiagnostică sunt următoarele (Anastasi, A., 1988, 1996):

1) Achiziţiile academice, evidenţiate prin note, medii, evidenţieri speciale, rezultate la concursuri, promovări, graduări, recompense, burse etc., sunt foarte frecvent folosite pentru validarea testelor de inteligenţă. De aceea, mai mulţi autori au considerat aceste teste ca fiind modalităţi de măsurare a aptitudinilor şcolare. De asemenea, achiziţiile academice sunt folosite şi pentru validarea testelor de măsurare a personalităţii şi a testelor pentru măsurarea aptitudinilor multiple;

2) Performanţa în cadrul unui domeniu de formare specializat este folosită în calitate de criteriu pentru validarea unor teste folosite pentru măsurarea aptitudinilor speciale. Pot fi date ca exemple: - achiziţiile finale în cadrul cursurilor de formare profesională specială sunt folosite pentru validarea testelor ce măsoară aptitudinile tehnice; - performanţa în anumite ramuri artistice pot fi folosite în vederea validării testelor pentru măsurarea aptitudinilor artistice;

3) Performanţa muncii (job performance) apare în calitate de criteriu folosit pentru validarea, în primul rând, a testelor pentru aptitudini speciale. Totodată, ea este utilizată, dar

Copyright © DEPARTAMENT ID 2008

în mai mică măsură, pentru validarea testelor de inteligență generală şi a testelor de personalitate;
4) Diagnoza psihiatrică este folosită în calitate de criteriu pentru evidențierea validității testelor de personalitate, dacă este bazată pe observarea prelungită şi detaliată a cazului;
5) Corelațiile dintre un test nou şi testele disponibile anterioare; de exemplu, validarea testelor de grup folosind bateria Stanford-Binet;
6) Metoda grupurilor contrastante, ce implică un criteriu compozit. De exemplu, validitatea unor teste de aptitudini muzicale sau a unor teste de aptitudini tehnice pot fi verificate prin compararea scorurilor obținute de studenții admişi în instituțiile de artă sau de inginerie cu scorurile obținute de cei respinşi la examenele de selecție.

Temă de reflecție/autoevaluare:
Dacă notele şcolare ridică multe probleme privind modul în care ele reflectă achizițiile reale ale elevilor, în ce măsură pot fi acceptate ca o formă de validare a unor teste de inteligență ?

3. Validitatea de construct (construct-related validation).

Această formă de validitate vizează măsura în care un test are capacitatea de a măsura un anumit construct teoretic. Fiecare test îşi propune să măsoare o anumită însuşire sau trăsătură psihică, cum ar fi, de exemplu: aptitudine şcolară, apreciere spațială, inteligență, comprehensiune verbală, fluență verbală, coordonare motorie, nevrozism, dominanță, anxietate, responsabilitate, intro-extroversie, timiditate etc. Fiecare dintre aceşti termini reprezintă o abstracție şi el reflectă o anumită parte a comportamentului uman. Pentru a-l putea măsura este necesar să inventariem cât mai multe dintre actele comportamentale care sunt observabile şi măsurabile. De exemplu, ce înseamnă că un individ este introvertit ? Alcătuim o listă de acte comportamentale, ceea ce înseamnă că operaționalizăm conceptual de introversie, care va cuprinde: - nu-i place să vorbească prea mult cu alții; - nu are prieteni; - stabileşte greu relații cu alte persoane; - nu răspunde dacă nu este întrebat; - nu-i place să atragă atenția altora asupra lui; - îi place să stea mai izolat; - îi place să-şi analizeze propriile gânduri; nu doreşte să meargă la petreceri etc.

Validarea de construct cuprinde mai multe etape, care sunt diferite de la un autor la altul, dar care în final atinge acelaşi obiectiv. Astfel, Cronbach, 1979 (apud Stan, A., 2002) ia în considerare 3 etape: 1. Lansarea ipotezei în privința constructului care elucidează comportamentul la test. Acesta este un act de reprezentare, care se bazează pe observarea comportamentului la test şi pe cercetarea logică a testului; 2. Deducția ipotezelor verificabile din teoriile localizate în construct. Aceasta este o operație pur logică; 3. Realizarea unei cercetări empirice pentru verificarea acestor ipoteze.

Alți autori, respectiv, Guthke, Böttcher şi Sprung, 1991 (apud Stan, A., 2002) consideră că demersul pentru validare este constituit din patru trepte şi anume:
1. Colectarea din fundamentele teoriilor asupra obiectului diagnosticului (de exemplu, anxietate, introversiune) a unei serii de aserțiuni, cum a r fi: a) despre relațiile presupus pozitive ale constructului implicat în test şi alte constructe; b) despre neconcordanțele presupuse de relația dintre constructele cercetate şi alte constructe; c) despre relația dintre constructele cercetate şi anumite variabile observate (moduri de comportament, produse ale activității etc.);

Copyright © DEPARTAMENT ID 2008

2. Alegerea, mai ales pentru dezvoltarea unor teste, a acelora care apar adaptate pentru constructele indicate la "a" şi "b";
3. Formularea unei serii de ipoteze care se referă la măsurarea constructelor corespunzătoare (tipul "a" se referă la validitatea convergentă, tipul "b" la cea discriminantă, iar tipul "c" la validarea criteriului;
4. Verificarea sau falsificarea, respectiv modificarea ipotezelor anterioare, pe baza cercetării.

Printre modalităţile de evidenţiere a validităţii constructului menţionăm următoarele:
1. Diferenţierea vârstei, criteriu folosit pentru validarea mai ales a testelor de inteligenţă. De exemplu, în cazul bateriei Stanford-Binet, este folosită vârsta cronologică pentru a se evidenţia dacă scorurile obţinute arată o creştere progresivă odată cu creşterea învârstă a subiectului;
2. Corelaţiile cu alte teste similare anterioare, care măsoară acelaşi tip de comportament. Un test nou de inteligenţă nou este declarat valid dacă el corelează cu un alt test de inteligenţă a cărui validitate a fost deja verificată;
3. Analiza factorială, ce ne apare în calitate de tehnică statistică rafinată pentru analizarea interrelaţiilor datelor comportamentale. Scopul major al analizei factoriale este de a simplifica descrierea comportamentului supus măsurării, reducând numărul mare de variabile luate în atenţie la câţiva *factori* sau *trăsături comune*.
4. Analiza consistenţei interne, folosită mai ales pentru categoria testelor de personalitate, în raport cu care criteriul nu este altceva decât scorul total al testului însuşi.
5. Matrice multi-trăsături - multi-metode (*multi-trait - multimethod matrix*).

Această ultimă modalitate o vom detalia mai mult, pentru că ea este prezentată practic în toate tratatele de psihodiagnostic (Anastasi, A., 1988, 1996) şi pentru că este una dintre cele mai utile. Este vorba despre un proiect experimental propus de către Campbell şi Fiske (1959). Ei diferenţiază două tipuri de validitate, respectiv, *validitatea convergentă* şi *validitatea discriminatorie*. Validitatea convergentă se referă la faptul că un test corelează la nivel înalt cu alte variabile cu care teoretic trebuie să coreleze. De exemplu, rezultatele unui elev la un test pentru măsurarea aptitudinilor matematice corelează cu rezultatele sale şcolare la matematică (validitate convergentă). Pe de altă parte, este de aşteptat ca performanţele la acelaşi test să nu coreleze cu rezultatele lui şcolare la literatură (validitate discriminatorie).

Revenind la proiectul experimental propus de Campbell şi Fiske, menţionăm că procedura folosită solicită măsurarea a două sau mai multe trăsături prin două sau mai multe metode. Concret, în proiectul lor: - cele trei trăsături sunt trăsături de personalitate, respectiv: A (dominanţa); B (sociabilitatea); C (motivaţia achiziţionării); - cele trei metode sunt: 1) inventar de personalitate; 2) test proiectiv; 3) clasificarea în perechi. Pot rezulta următoarele combinaţii posibile:
A_1, A_2, A_3 = dominanţa la testele 1, 2, 3;
B_1, B_2, B_3 = sociabilitatea la testele 1, 2, 3;
C_1, C_2, C_3 = motivaţia achiziţionării la testele 1, 2, 3.

Se alcătuieşte un table cu dublă intrare, în cadrul căruia regăsim următoarele tipuri de informaţii:
- valoarea coeficienţilor de fidelitate;
- valoarea coeficienţilor de validitate, rezultaţi în urma corelării scorurilor obţinute pentru aceeaşi trăsătură prin diferite metode;
- corelaţiile dintre diferite trăsături măsurate cu aceeaşi metodă;

Copyright © DEPARTAMENT ID 2008

- corelaţiile dintre diferite trăsături măsurate cu diferite metode.

În urma analizei datelor din table au rezultat următoarele concluzii:

a) pentru ca validitatea constructului să fie satisfăcătoare, valoarea coeficienţilor de validitate trebuie să fie mai mare decât valoarea corelaţiilor dintre trăsturile diferite măsurate cu metode diferite;

b) valoarea coeficienţilor de validitate trebuie să fie mai mare decât valoarea corelaţiilor dintre trăsături diferite măsurate cu aceeaşi metodă.

Şi acum întrebarea: cât de mare poate fi valoarea coeficientului de validitate ? Nu există un răspuns acceptat unanim de către autori. Desigur, ca o cerinţă generală, cu cât valoarea este mai mare cu atât mai valid este testul respectiv. Să nu uităm însă, că absolut întotdeauna valoarea coeficientului de validitate al unui test este mai mică decât valoarea coeficientului de fidelitate al aceluiaşi test.

Unii autori (Anastasi, A., 1988, 1996) susţin că valoarea coeficientului validităţii trebuie să fie destul de mare pentru a fi statistic semnificativă la un nivel acceptabil, cum ar fi 0.01 sau 0.05. Alţi autori, cum este cazul lui M. Smith (apud Stan, A., 2002), propun o gradare a semnificaţiilor unui indice de validitate:

a) peste 0,50 - excelent;
b) între 0,40 şi 0,49 - bun;
c) între 0,30 şi 0,39 - acceptabil;
d) sub 0,30 - slab.

A. Stan atenţionează însă că un coeficient de validitate mai mare de 0,70 este excepţional şi că el merită o examinare atentă când este raportat.

Temă de reflecţie/autoevaluare:

1. Încercaţi să operaţionalizaţi constructele: timiditate şi anxietate
2. De ce întotdeauna valoarea coeficientului de validitate al unui test este mai mică decât valoarea coeficientului de fidelitate ?

Întrebări de autoevaluare:

1. De ce validitatea de conţinut nu este valabilă pentru testele de personalitate şi testele de aptitudini ?
2. În afară de testele de cunoştinţe pentru ce categorii de teste mai este valabilă validitatea de conţinut ?
3. Ce cuprind specificaţiile testului ?
4. La ce ne ajută formula lui Lawshe ?
5. Care este specificul validităţii de criteriu ?
6. Care este deosebirea dintre predictori şi criteriu ?
7. Prin ce se deosebeşte validitatea concurentă de validitatea predictivă ?
8. În ce constă fenomenul de contaminare a criteriului şi cum poate fi el contracarat ?
9. Care sunt cele mai cunoscute criterii utilizate în practica psihodiagnostică ?
10. În ce constă validitatea privitoare la construct ?
11. Care este poziţia diferiţilor autori privind etapele parcurse pentru evidenţierea validităţii de construct ?
12. Care sunt principalele modalităţi de evidenţiere a validităţii de construct ?
13. Prin ce se deosebeşte validitatea convergentă de validitatea discriminatorie ?
14. În ce constă şi la ce foloseşte proiectul experimental propus de către Campbell şi Fiske ?
15. Ce putem spune despre valoarea coeficientului de validitate ?

Copyright © DEPARTAMENT ID 2008

16. De ce întotdeauna valoarea coeficientului de validitate este mai mică decât valoarea coeficientului de fidelitate ?

BIBLIOGRAFIE MINIMALĂ

1. ANASTASI, A. (1988, 1996). *Psychological testing*. New York: Macmillan.
2. COHEN, R.J., SWERDLIK, M.E. (2005). *Psychological Testing and Assessment*. *An Introduction to Tests and Assessment.* McGraw-Hill International Edition.
2. CLINCIU, A.I. (2005). *Psihodiagnostic.* Braşov: Editura Universităţii "Transilvania".
3. MINULESCU, M. (2003). Teorie şi practică în psihodiagnoză. Bucureşti: Editura Fundaţiei România de Mâine.
4. STAN, A. (2002). Testul psihologic. Evoluţie, construcţie, aplicaţie. Iaşi: Polirom.

Copyright © DEPARTAMENT ID 2008

Unitatea de învăţare 4
Testul psihologic şi selecţia profesională

Cuprins:
1. Eroarea standard a estimării
2. Criterii de selectare a testelor
3. Locul şi rolul testului în selecţia profesională

Obiective:
La sfârşitul acestei unităţi de învăţare, studenţii vor fi capabili să:
- cunoască în ce constă eroarea standard a estimării;
- poată să dea exemple operând cu SEE;
- identifice criteriile în baza cărora selectăm testele pentru a le folosi în practică;
- folosească tabelele Taylor-Russell;
- determine capacitatea de selecţie a unor teste.

1.

Eroarea standard a estimării (*Standard Error of Estimation* – SEE).

Este foarte asemănătoare cu eroarea standard a măsurării (SEM), numai că aceasta nu mai are legătură cu fidelitatea, ci cu validitatea. SEM ajută la determinarea mărimii erorii ce poate fi aşteptată ca urmare a faptului că scorul unui subiect este rezultatul nonfidelităţii testului. La rândul ei, SEE ajută la determinarea erorii ce poate fi aşteptată privitor la scorul prognozat al subiectului la un criteriu, ca rezultat al validităţii imperfecte a testului. Aşa după cum deja cunoaştem, în baza rezultatelor obţinute la un test folosit pentru selectarea subiecţilor, anticipăm poziţia sau chiar scorul unui subiect la criteriu. În cazul în care testul ar fi perfect valid, de câte ori am aplica testul am anticipa exact acelaşi scor la criteriu, ceea ce în realitate este imposibil. Datorită valorii relativ reduse a validităţii testului (vă amintiţi că valoarea coeficientului de validitate este întotdeauna mai mică decât valoarea coeficientului de fidelitate ?), poziţia anticipată a subiectului la criteriu va oscila între anumite limite, care pot fi precizate cu ajutorul lui SEE. Formula de calcul este următoarea:

$$SEE = SD_y \sqrt{1 - r_{xy}^2}$$

în care: SD_y = deviaţia standard la scorurile la criteriu;
r_{xy}^2 = pătratul coeficientului de validitate

Să vedem ce se întâmplă în cazul în care valoarea coeficientului de validitate = 1.00, adică testul ar fi perfect valid. Înlocuim în formulă şi vom constata că valoarea lui SEE ar fi zero, deci precizia anticipării scorului subiectului la criteriu ar fi maximă. În cazul în care, însă, coeficientul de validitate ar avea valoarea zero, înlocuind în formulă, vom constata că mărimea erorii standard a estimării ar fi atât cât este deviaţia standard. Deci, predicţia nu ar mai avea nicio precizie, ea fiind total întâmplătoare. Inutil să mai spunem că valoarea predictivă a testului ar fi practic nulă.

Desigur, calcularea valorii lui SEE intră în obligaţia constructorului testului iar informaţiile privind acest tip de eroare trebuie să fie incluse în manualul testului.

Copyright © DEPARTAMENT ID 2008

Cunosc\nd valoarea lui SEE noi putem să anticipăm între ce limite se va încadra scorul subiectului prognozat la criteriu. Să presupunem că scorul prognozat este 100. Dacă reaplicăm testul de 100 de ori, scorul prognozat va oscila în jurul scorului egal cu 100, respectiv, jumătate dintre ele vor fi mai mari şi jumătate vor fi mai mici. Apelând la curba lui Gauss, vom fi puşi în faţa unei distribuţii normale, având media = 100. SEE va reprezenta deviaţia standard a acestei distribuţii. Drept urmare: 34% din scorurile subiectului vor cădea între 100 şi + 1 SEE (însemnând media şi plus o deviaţie standard) şi alte 34% din scoruri vor cădea între 100 şi - 1 SEE (însemnând media şi minus o deviaţie standard). Luate împreună, 68% din scorurile subiectului vor cădea în intervalul definit de 100 ± 1 SEE. Deci, dacă valoarea lui SEE ar fi 10, noi putem fi siguri că, în caz de reaplicare a testului în vederea anticipării scorului la criteriu, acesta din urmă va cădea în intervalul 90 şi 110, la un nivel de încredere de 68%. La un nivel de încredere de 95%, scorul prognozat ar cădea în intervalul 100 ± 2 SEE, respectiv, pentru exemplul dat, în intervalul 80 şi 120. La un nivel de încredere de 99%, scorul prognozat ar cădea în intervalul 100 ± 3 SEE, respectiv, pentru exemplul nostru, în intervalul 70 şi 130.

Exact ca şi în cazul erorii standard a măsurării, formula pentru construirea intervalului de încredere este următoarea:

X ± (scorul z) (SEE), în care:

X + (scorul z) (SEE) = limita superioară a intervalului

X - (scorul z) (SEE) = limita inferioară a intervalului.

Aşa după cum am precizat de mai multe ori, valoarea coeficientului de validitate al testelor este relativ mică, ceea ce înseamnă că, de regulă, eroarea scorurilor prognozate este considerabilă. De aceea, în realitate, cele mai multe teste sunt folosite nu pentru a prevedea exact scorul unui subiect la criteriu, ci mai mult pentru a determina dacă el va depăşi un standard minim al performanţei sau o linie de demarcaţie între succes-eşec (Anastasi, A., 1988, 1996).

Temă de reflecţie/autoevaluare:

1. Ce este eroarea standard a estimării (SEE)şi care este relaţia sa cu validitatea testului ?
2. Dacă valoarea lui SEE este 5 şi scorul prognozat al subiectului în baza rezultatelor la un test aplicat este 120, între ce limite se va întinde scorul prognozat, la un nivel de încredere de 99%, în caz de reaplicare a testului ?

3. **Criterii de selectare a testelor.**

În practica psihgodiagnostică există foarte multe teste, de aceea, se pune problema selectării lor, atât în funcţie de caracteristicile psihometrice, cât şi în funcţie de scopul acţiunii de măsurare. Desigur, trebuie evitată pe cât posibil utilizarea unui singur test, mai ales atunci când trebuie luată o decizie importantă pentru subiect. Unele teste sunt deja foarte cunoscute, ele fiind aplicate de foarte mulţi psihologi. În unele ţări, cum este cazul S.U.A. s-au realizat chiar ierarhizări ale testelor în funcţie de preferinţa manifestată de practicieni. Teste precum MMPI, CPI, WAIS, WISC, WPPSI, Rorschach, TAT etc. s-au impus mai demult în partea de început a acestor ierarhizări. Cu toate acestea, selectarea testelor pe care trebuie să le aplice psihologul nu este o sarcină uşoară, el trebuind să dovedească multă flexibilitate şi multă răspundere în acest sens.

Copyright © DEPARTAMENT ID 2008

Printre factorii care trebuie să fie luaţi în consideraţie în selectarea testelor pot fi incluşi următorii:

- motivul pentru care subiectul a fost trimis la psiholog pentru măsurarea şi evaluarea psihologică;
- vârsta subiectului;
- probleme prezentate de subiect de ordin snzorial (văz, auz) sau de ordin motor;
- perioada de timp disponibilă pentru testare;
- disponibilitatea anumitor teste;
- abilitatea subiectului de a citi (pentru cazurile în care vom folosi teste de personalitate sau anumite inventare de interese vocaţionale);
- mediul socio-cultural din care provine subiectul.

În continuare vom prezenta câteva exemple (Bishop, E., 1992) de modul în care se pot combina testele, ţinându-se seamă de scopul trimiterii spre examinare şi de vârsta subiecţilor.

Exemplul 1: pentru un adult care este trimis pentru o examinare psihologică completă în vederea stabilirii exacte a diagnosticului şi planificării tratamentului, fiind suspectat că are probleme psihiatrice, cum ar fi, depresia sau anxietatea se poate folosi următoarea baterie de teste. Testele sunt prezentate în ordinea aplicării lor:

- Scala de inteligenţă pentru adulţi a lui Wechsler (WAIS – III);
- Inventarul multifazic de personalitate Minnesota (MMPI – 2);
- Testul Bender Gestalt;
- Testul tematic de apercepţie (TAT);
- Testul completării de fraze (Incomplete Sentence Blank).

Acest grup de teste, atât obiective, cât şi proiective vor oferi informaţii privind abilităţile cognitive şi funcţionarea personalităţii.

Exemplul 2: pentru copiii ce urmează a fi evaluaţi pentru plasarea lor într-o instituţie pentru educaţie specială, se poate folosi următoarea baterie de teste:

- Scala de inteligenţă a lui Wechsler pentru copii (WISC-III);
- Testul pentru achiziţii şcolare Woodcock-Johnson (Woodcock-Johnson Achievement Test – Revised);
- Testul Bender-Gestalt;
- Testul de dezvoltare a integrării vizual-motorii (The Developmental Test of Visual-Motor Integration);
- The Draw a Person (DAP) sau House-Tree-Person (HTP).

Deci această baterie vizează mai multe aspecte, respectiv, nivelul de dezvoltare intelectuală, nivelul şi tipurile de achiziţii şcolare, nivelul de funcţionare cognitivă, nivelul de funcţionare a personalităţii. Dacă este necesară confirmarea unui diagnostic posibil privind retardarea mentală, trebuie să fie adăugate alte instrumente de măsurare a comportamentului adaptativ, cum este cazul Scalelor Vineland pentru Comportamentul Adaptativ (Vineland Adaptative Behavior Scales). Aceste scale sunt foarte necesare deoarece diagnoza retardării mentale presupune evidenţierea unor întârzieri semnificative atât în ceea ce priveşte funcţionarea cognitivă, cât şi în ceea ce priveşte comportamentul adaptativ.

Exemplul 3. Dacă un adolescent este evaluat pentru posibile servicii de reabilitare vocaţională, bateria folosită ar trebui să cuprindă: teste pentru abilităţi cognitive, teste de

Copyright © DEPARTAMENT ID 2008

achiziţii şcolare, teste pentru planificare şi organizare, teste pentru interese vocaţionale şi teste de personalitate. Concret, această baterie ar cuprinde:
- Scala de inteligenţă a lui Wechsler pentru adulţi (WAIS-III);
- Wide Range Achievement Test- Revised;
- Strong Vocational Interst Blank;
- Incomplete Sentences Blank.

Exemplul 4. Pentru evaluarea unui copil mic şi foarte mic, în vederea beneficierii de un suport financiar datorită unor dizabilităţi, se poate utiliza o baterie care să cuprindă:
- Testele Bayley ale dezvoltării copilului (Bayley Tests of Infant Development;
- Scalele Vineland pentru comportamentul adaptativ (The Vineland Adaptive behavior Scales;
- Observarea interacţiunii mamă-copil.

O grijă deosebită trebuie acordată selectării testelor potrivite pentru evaluările judiciare (forensic evaluations).

Exemplul 5. În cazul stabilirii custodiei pentru un copil, testele de evaluare sunt aplicate şi pe copii şi pe părinţi, la care se adaugă interviul clinic şi observarea interacţiunii părinte-copil. Ca teste pentru copii pot fi folosite următoarele:
- Testul Aperceptiv Robert pentru Copii (Roberts Apperception for Children);
- Testul vocabular în imagini Peabody (Peabody Picture Vocabulary Test- Revised);
- Desenul kinetic al familiei (Kinetic Family Drawing);
- Completare de propoziţii (Incomplete Sentences Blank).

Pentru adulţi se aplică testele:
- Inventarul multifazic de personalitate Minnesota (MMPI-2);
- Completare de propoziţii (Incomplete Sentences Blank);
- Indexul stresului parental (Parenting Stress Index).

Teme de reflecţie/autoevaluare:
1. De ce credeţi că este necesar, precum în exemplele de mai sus, să fie incluse într-o baterie de testare mai multe teste ?
2. Care sunt cele mai importante criterii ce se iau în atenţie în selectarea unor teste folosite în acţiunile de măsurare psihologică ?

3. Locul şi rolul testului în selecţia profesională.

Testul este foarte frecvent folosit în acţiunile de selecţie profesională. În urma aplicării lui rezultă două categorii de subiecţi: a) cei care reuşesc la test; b) cei care nu reuşesc la test. Primii mai sunt denumiţi subiecţi *admişi* iar ceilalţi sunt denumiţi subiecţi *respinşi*. Cei care organizează examenul de selecţie trebuie să stabilească linia de demarcaţie (*cut-off score*) dintre subiecţii admişi şi cei respinşi. Această linie poate fi foarte apropiată de partea de început a listei candidaţilor (de exemplu, dacă se pot obţine maximum de 100 de puncte la test, vor fi declaraţi reuşiţi numai cei ce obţin peste 95 de puncte) sau, dimpotrivă, poate fi foarte mult îndepărtată de partea de început (în cadrul exemplului dat pot fi declaraţi admişi cei ce obţin cel puţin 25 de puncte). Este clar că, în primul caz, interesul organizatorului examenului este de a face o selecţie foarte severă, iar în al doilea caz, interesul este de a accepta cât mai mulţi candidaţi.

Dacă testul psihologic folosit pentru selecţie ar fi un instrument perfect (coeficientul de fidelitate = 1,00 şi coeficientul de validitate = 1,00) atunci, în cadrul acţiunii de selecţie, ar fi sigur departajaţi subiecţii buni (admişi) de cei slabi (respinşi). Deci nu ar exista nicio greşeală, în sensul că pot fi admişi candidaţi slabi şi respinşi candidaţi buni.

Copyright © DEPARTAMENT ID 2008

Cum însă, în realitate, testul psihologic nu este un instrument de măsurare perfect (niciodată valoarea coeficientului de fidelitate şi a coeficientului de validitate nu este 1,00), el poate „greşi", în sensul că, pe lângă subiecţii admişi pe „drept" pot fi incluşi şi subiecţi care ar trebui să fie respinşi. De asemenea, în categoria subiecţilor respinşi pe „drept", el poate include şi unii subiecţi, care ar trebui să fie, de fapt, admişi.

Iată, aşadar, că, în realitate, în urma examenului de selecţie rezultă nu doar două categorii, ci patru: admişi, respinşi, fals admişi, fals respinşi. Desigur, ultimile două categorii ridică probleme serioase în cazul în care procentul celor ce le aparţin este mare. Falşii admişi sunt cei care nu au capacităţi şi aptitudini pentru activitatea în raport cu care s-a făcut selecţia şi menţinerea lor în acţiunile de formare (training) vor necesita costuri mari inutile, deoarece ei oricum vor eşua când vor fi evaluaţi la criteriu. Falşii respinşi înseamnă, candidaţi buni „pierduţi", care, dacă ar fi fost admişi, ar fi reuşit sigur când ar fi fost evaluaţi la criteriu.

Cum ar trebui să procedeze organizatorul examenului pentru a reduce numarul falşilor pozitivi (cei admişi pe nedrept) ? O măsură simplă ar fi să ridice linia de demarcaţie, scorul de departajare (cut-off score) cât mai aproape de partea de început a listei candidadaţilor. Cei care obţin scoruri foarte mari este greu de presupus că ei vor aparţine grupului falşilor pozitivi. Şi invers, dacă organizatorul examenului de selecţie nu doreşte să-i piardă pe cei care au capacităţile şi aptitudinile măsurate de test (să nu-i piardă pe falşii negativi), poate coborî linia de demercaţie (cut-off score) cât mai jos posibil.

Iată, aşadar, că se pune problema în ce măsură testul psihologic folosit în selecţia candidaţilor are capacitatea de a face o predicţie corectă. Mai întâi, vorbim despre *rata de selecţie* a cestuia (*hit rate*), ce se referă la proporţia cazurilor în care el anticipă corect succesul sau eşecul. Pe de altă parte, având în vedere rezultatele la criteriu, se poate determina *rata de bază* (*base rate*), care este proporţia celor care reuşesc şi a celor care nu reuşesc la criteriu.

Să luăm şi noi un exemplu concret (Kaplan, M.R., Saccuzzo, P. Dennis, 1993). Este vorba despre un test cu 83% precizie predictivă şi 80% precizie a detectării.

		Rezultate la testul psihologic		
		Afecţiune neurologică	Normal	Total
Actual	Afecţiune neurologică	A 8	B 2	10
	Normal	C 15	D 75	90
	Total	23	77	100

Citind datele din tabel, rezultă că, în urma testării psihologice a 100 de subiecţi, a rezultat că 23 au afecţiune neurologică şi că 77 dintre ei nu au o asemenea afecţiune. Pe de altă parte, tot din tabel reiese că, în urma efectării diagnosticului medical, a reieşit că 10 subiecţi au afecţiune neurologică iar ceilalţi 90 sunt normali. Ce rezultă de aici ?

În tabel există două tipuri de preziceri corecte: a) din 10 subiecţi cu afecţiune neurologică, testul detectează 8, ceea ce înseamnă că rata de detecţie este de 80%; b) testul arată că 75 dintre subiecţi sunt normali, fiind în total acord cu rezultatul examenului medical. Deci, din 100 de subiecţi, testul stabileşte corect „diagnosticul" pentru 83 dintre ei. Putem spune, în consecinţă, că acest instrument psihodiagnostic prezintă un procent de 83%

Copyright © DEPARTAMENT ID 2008

a acurateței sau *preciziei predictive*. Totodată, însă, există şi două tipuri de eşecuri în predicție. Astfel, există două cazuri considerate, în baza testului, ca neavând afecțiune neurologică, în realitate nefiind aşa. De asemenea, conform datelor obținute la test, 15 cazuri prezintă o afecțiune neurologică, situație neconfirmată de examenul medical.

Aşadar, din tabelul de mai sus reies următoarele tipuri de informații:

- A şi D = predicție corectă a testului
- B = fals negativi
- C = fals pozitivi
- A/(A + B) = rata de detecție
- D/(C + D) = specificitate
- (A + D)/(A + B + C + D) = rata acurateței sau preciziei predictive

Concluzia principală este aceea că acest test este relativ bun în detectarea afecțiunilor neurologice, deoarece 10% din toți subiecții au asemenea afecțiuni iar testul detectează 80% din cazuri.

Pentru a cunoaşte cât de mult ne poate ajuta un test psihologic în acțiunile de selecție, Taylor şi Russell au oferit o metodă pentru evaluarea validității în relație cu cantitatea de informație cu care testul contribuie dincolo de *rata de bază*. Această metodă a măsurării valorii testelor este prezentată într-o serie de tabele, cunoscute sub numele de *tabele Taylor-Russell* (Kaplan, M.R., Saccuzzo, P.D., 1993). Cel care doreşte să utilizeze aceste tabele trebuie să dispună de următoarele informații:

1. Definirea succesului şi precizarea cât mai clară a limtei de demarcație între cele două categorii de subiecți (*admişi-respinşi*);
2. Determinarea ratei de bază, adică procentul persoanelor care vor reuşi în condițiile în care nu se foloseşte niciun test;
3. Definirea ratei de selecție: procentul candidaților selectați sau admişi;
4. Determinarea coeficientului de validitate. Acesta, de regulă, vizează corelația dintre rezultatele la test şi rezultatele la criteriu.

Tabelele Taylor-Russell oferă probabilitatea ca o persoană selectată pe baza scorului la test să reuşească la criteriu. Există câte un tabel diferit pentru fiecare rată de bază. Tabelul care este cel mai cunoscut şi care poate fi regăsit apropate în toate tratatele de psihodiagnostic este cel cu o rată de bază: 0,60.

Pentru a utiliza tabelul se ia în atenție şirul ce cuprinde valorile validității testului ce urmează a fi utilizat pentru selecție. Apoi se găseşte coloana asociată procentului celor ce vor fi selectați. Numărul găsit la intersectarea şirului şi coloanei ne oferă o estimare a celor ce vor reuşi în cazul în care vor fi selectați pe bază de test.

Iată un exemplu concret (Kaplan, M.R., Saccuzzo, P.D., 1993). Să presupunem că avem în calitate de criteriu notele (mediile) obținute de elevi după terminarea unui an şcolar. Având în vedere că limita de demarcație (cutt-off score) este 5,00, 60% dintre ei sunt declarați reuşiți. Deci rata de bază = 60%. Apoi, considerăm că vom utiliza pentru selecția elevilor un test de inteligență. Rata de selecție va fi 60%, deoarece avem numai atâtea locuri. Valoarea coeficientului de corelație dintre acest test şi criteriu este deja cunoscută: 0,30. Pentru a estima câți candidați vor reuşi dacă ei sunt selectați în baza testului de inteligență folosim tabelul Taylor-Russell. Vom găsi procentul de .66 sau 66%. Deci, diferența este doar de 6% față de selecția bazată pe întâmplare. Dacă valoarea coeficientului de validitate ar fi fost 0,50, atunci procentul ar fi fost de 73%, ceea ce înseamnă o diferență de 13% față de selecția bazată pe întâmplare.

Copyright © DEPARTAMENT ID 2008

Tabelul Taylor-Russell pentru rata de bază = 60%

	Rata de selecție										
Validitatea	**0.5**	**0,10**	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**	**0.95**
.00	.60	.60	.60	.60	.60	.60	.60	.60	.60	.60	.60
.05	.64	.63	.63	.62	.62	.62	.61	.61	.61	.60	.60
.10	.68	.67	.65	.64	.64	.63	.63	.62	.61	.61	.60
.15	.71	.70	.68	.67	.66	.65	.64	.63	.62	.61	.61
.20	.75	.73	.71	.69	.67	.66	.65	.64	.63	.62	.61
.25	.78	.76	.73	.71	.69	.68	.66	.65	.63	.62	.61
.30	.82	.79	.76	.73	.71	.69	.68	.66	.64	.62	.61
.35	.85	.82	.78	.75	.73	.71	.69	.67	.65	.63	.62
.40	.88	.85	.81	.78	.75	.73	.70	.68	.66	.63	.62
.45	.90	.87	.83	.80	.77	.74	.72	.69	.66	.64	.62
.50	.93	.90	.86	.82	.79	.76	.73	.70	.67	.64	.62
.55	.95	.92	.88	.84	.81	.78	.75	.71	.68	.64	.62
.60	.96	.94	.90	.87	.83	.80	.76	.73	.69	.65	.63
.65	.98	.96	.92	.89	.85	.82	.78	.74	.70	.65	.63
.70	.99	.97	.94	.91	.87	.84	.80	.75	.71	.66	.63
.75	.99	.99	.96	.93	.90	.86	.81	.77	.71	.66	.63
.80	1.00	.99	.98	.95	.92	.88	.83	.78	.72	.66	.63
.85	1.00	1.00	.99	.97	.95	.91	.86	.80	.73	.66	.63
.90	1.00	1.00	1.00	.99	.97	.94	.88	.82	.74	.67	.63
.95	1.00	1.00	1.00	1.00	.99	.97	.92	.84	.75	.67	.63
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.86	.75	.67	.63

După cum am afirmat mai înainte, în urma examenului de selecție, este posibil ca unii candidați, deşi valoroşi, să nu fie selectați. De aceea, se ridică problema existenței unei modalități de estimare a proporției candidaților declarați admişi care vor confirma la criteriu şi procentul celor care ar fi reuşit la criteriu dacă ar fi fost selectatți.

De exemplu (Kaplan, M.R., Saccuzzo, P.D.), să presupunem că un şef de personal la o companie doreşte să selecteze 30 de subiecți dintr-un număr de 100 de candidați. Se va folosi un test cu validitatea = .70. Rata de bază este 60%. Utilizând tabelul Taylor-Russell pentru o rată de bază de 60%, vom găsi procentul de .91, ceea ce înseamnă că 91% din candidații selectați vor reuşi la criteriu (91% x 30 = 27,3, deci, 27. Apoi, făcând diferența (30 – 27 = 3) rezultă că 3 candidați nu vor reuşi la criteriu.

Dacă şeful de personal s-a hotărât să angajeze 30 de candidați din 100 prezentatți la concurs, înseamnă că nu va angaja 70 dintre ei. Asta nu înseamnă, însă, că toți cei respinşi, adică toți cei 70 de candidați, dacă ar fi fost angajați, ar fi eşuat la criteriu. Situația este reflectată de tabelul de mai jos.

	Test	
	Admişi	Respinşi	Total

Copyright © DEPARTAMENT ID 2008

Criteriu	Succes	27	33	60
	Eşec	3	37	40
	Total	30	70	100

Ce se poate observa din acest tabel ? Din cei 60 de candidaţi care ar reuşi la criteriu sunt selectaţi, cu ajutorul testului, doar 27, ceea ce înseamnă că 33 de candidaţi, care ar reuşi la criteriu, nu sunt selectaţi. Totuşi, din cei 40 de candidaţi, care ar eşua la criteriu, 37 au fost bine diagnosticaţi. În concluzie, 90% (27/30) dintre candidaţii selectaţi cu ajutorul testului vor reuşi la criteriu iar 47% (33/70) dintre cei respinşi ar fi reuşit la criteriu. Procedura poate fi declarată a fi bună, deoarece procentul celor ce vor reuşi dintre cei selectaţi este mult mai mare.

Teme de reflecţie/autoevaluare:
1. Analizând tabelul Taylor-Russell, ce putem spune despre testele ce au validitatea mai mare iar rata de selecţie este mai mică ? Dar despre testele care au validitatea mai mică iar rata de selecţie este mai mare ?

Intrebări de autoevaluare:

1. Ce reprezintă eroarea standard a estimării (SEE) ?
2. Care este formula de calcul a erorii standard a estimării ?
3. Ce asemănări şi ce deosebiri sunt între eroarea standard a măsurării (SEM) şi eroarea standard a estimării (SEE) ?
4. Cine se ocupă de calcularea celor două forme de erori ?
5. Daţi cel puţin 2 exemple concrete privind utilizarea valorii lui SEE.
6. Care sunt cele mai importante criterii folosite în selectarea testelor ?
7. Daţi câteva exemple de modul în care se pot combina testele într-o baterie de teste.
8. De ce trebuie să includem mai multe teste în acţiunile de măsurare şi evaluare psihologică ?
9. Când folosim teste psihologice pentru selecţia de personal, câte categorii reale de candidaţi avem ?
10. Ce înseamnă rata de detecţie a testelor ?
11. Ce înseamnă rata acurateţei sau a preciziei predictive a testelor ?
12. Care sunt paşii parcurşi când folosim tabelele Taylor-Russell ?

BIBLIOGRAFIE

1. BISHOP, E. (1992). *Selection of psychological tests for different purposes:* children, adolescents, and adults. 50th International Convention of ICP, Olanda.
2. COHEN, J.R., SEWRDLIK; E:M: (2005). *Psychological Testing and Assessment. An Introduction to Tests and Measurement*. McGraw-Hill International Edition.
3. CLINCIU, A. (2005). *Psihodiagnostic*. Braşov: Ed. Universităţii "Transilvania".
4. KAPLAN, M. R., SACCUZZO, P.D. (1993). *Psychological Testing.* Brooks/Cole Publishing Company.
5. STAN, A. (2003). *Testul psihologic. Evoluţie, construcţie, aplicaţii*. Iaşi: Polirom.

Copyright © DEPARTAMENT ID 2008

MODULUL IV

ANALIZA ITEMULUI

Unitatea de învățare 1: Ce este analiza itemului ?

Cuprins:

1. Relația itemului cu testul

2. Tipuri de analiză a itemului

Obiective

La sfârşitul acestei unități de învățare, studenții vor fi capabili să:

- definească itemul ca element component al testului;
- cunoască relația dintre itemi şi testul luat ca întreg;
- cunoască funcțiile îndeplinite de item în cadrul testului;
- evalueze diferite tipuri de analize a itemului

1. Relația itemului cu testul

Am vorbit destul de mult despre caracteristicile psihometrice ale testelor, respectiv, despre fidelitate şi validitate. Acum vom vedea în ce măsură acestea depind de caracteristicile elementelor componente ale testelor, mai exact ale itemilor. Testul cuprinde un anumit număr de itemi şi urmăreşte să măsoare o anumită însuşire sau o anumită caracteristică psihică. Obiectivul va fi atins, însă, dacă fiecare dintre itemi va urmări şi va avea „puterea" să măsoare aceeaşi însuşire sau caracteristică psihică.

Sunt şi autori care nu folosesc noțiunea de item (apud Stan, A., 2002). De exemplu, David Magnusson, 1975, foloseşte noțiunile de *sarcini* şi *analiză de sarcini* în loc de *itemi* şi analiză de itemi. Gustav Lienert (1967) foloseşte, în cadrul tratatului său o singură dată noțiunea de item.

Să presupunem că, în cadrul unui test care conține 25 de itemi, la 5 dintre ei nu răspunde corect niciun subiect iar la alți 5 dintre ei, răspund corect toți subiecții. Este clar că cei 10 itemi, deşi sunt incluşi în structura testului, sunt inutili, în sensul că ei nu contribuie cu nimic la scoaterea în evidență a diferențelor interindividuale. Şi dacă aceşti itemi sunt „neputincioşi", atunci lungimea testului (numărul real de itemi) este, în realitate, mai mică cu 10 itemi. Şi dacă reducem lungimea testului (vă mai aduceți aminte ce se întâmplă ?) va scădea automat valoarea fidelității, cât şi a validității. În cazul exemplului nostru este bine ca itemii foarte dificili şi cei foarte uşori să fie eliminați din structura testului.

Dar ce este, de fapt, un item ? Cum îl putem defini ? În ce relație se află el cu testul luat ca întreg ? Vom oferi răspunsul la aceste întrebări, începând cu o definiție. Şi am preferat definiția dată de M. Reuchlin, în lucrarea *Grand dictionnaire de la psychologie* (1992):

Copyright © DEPARTAMENT ID 2008

„itemul este un element al unui test, constituind o situație particulară şi jucând rolul unui stimul la care subiectul nu poate da decât un răspuns dintre două sau mai multe răspunsuri teoretice prevăzute de test". Ce reiese din această definiție ? În primul rând, faptul că itemul este o secvență, o componentă a testului, un fel de „mintest" care are o anumită individualitate. În al doilea rând, faptul că itemul apare ca un stimul pentru subiect. În sfârşit, în al treilea rând, faptul că subiectul oferă un răspuns, care poate fi corect sau nu.

Desigur, legat de formele de răspuns, lucrurile se prezintă diferențiat în funcție de categoria testelor. De exemplu, la testele de performanță, răspunsul la item poate fi corect (subiectului i se acordă 1 punct) sau incorect (subiectului nu i se acordă niciun punct sau subiectului i se acordă 0 puncte). În cazul testelor de personalitate, dacă se folosesc scale cu mai multe variante de răspuns, subiectul poate obține un număr variabil de puncte, adică obține atâtea puncte câte sunt acordate treptei de răspuns pe care a ales-o subiectul.

Pentru a fi considerată item o componentă a testului trebuie să aibă individualitate informațională pertinentă, adică să contribuie distinct, clar şi conturat, la scorul total al testului (Stan, A., 2002). Un exemplu oferit de autor, care nu constituie item, este bararea unei litere, O sau C, la testul de atenție Bourdon-Amfimov.

S-a pus şi întrebarea: câți itemi trebuie să aibă un test ? Privind spre practică, putem întâlni şi teste cu mai puțini itemi şi teste cu foarte mulți itemi, cum este cazul chestionarelor şi inventarelor de personalitate care au sute de itemi. Noi ne reamintim faptul că un test este de aşteptat să fie mai fidel şi mai valid cu cât are mai mulți itemi. Referitor la cel mai mic număr de itemi pe care poate să-l aibă un anumit test, Paul Kline, o autoritate în domeniul psihodiagnosticului, susține că acesta nu poate fi mai mic de 10. Deci, niciun test şi niciun chestionar nu poate fi acceptat în calitate de instrument psihodiagnostic dacă nu are cel puțin 10 itemi.

Un alt aspect important este cel legat de modul în care se formulează itemii. Sarcina nu este deloc uşoară iar unii autori (DeVellis, 1991) a elaborat un fel de ghid ce cuprinde mai multe recomandări din care Kaplan, K.R. şi Saccuzzo, D. (2005) rețin următoarele şase:

1. Definiți clar ceea ce vreți să măsurați. Pentru aceasta utilizați teoria substantivului în calitate de ghid şi încercați să faceți itemii atât de specifici cât este posibil;
2. Creați o bancă de itemi. Teoretic, toți itemii sunt aleşi întâmplător dintr-un univers de conținut al itemilor. În practică, totuşi, grija în selectarea şi dezvoltarea itemilor este valabilă. Evitați itemii redundanți. În fazele inițiale poate doriți să scrieți 3 sau 4 variante pentru fiecare item ce urmează a fi utilizat în structura noului test;
3. Evitați, pe cât posibil, itemii cu lungime mare, deoarece aceştia sunt buni foarte rar;
4. Mențineți nivelul de dificultate al lecturii şi înțelegerii cât mai potrivit pentru cei care vor răspunde la test;
5. Evitați itemii cu un caracter echivoc, adică cei care conduc la două sau mai multe idei în acelaşi timp. De exemplu, să ne referim la un item care solicită respondentului să răspundă cu „acord" sau „dezacord" la afirmația „Votez democrații deoarece sprijin programe sociale". Aici sunt, de fapt, două afirmații în raport cu care respondentul trebuie să-şi exprime acordul: „Votez democrații" şi „Eu sprijin programele sociale";
6. Combinați itemii formulați pozitiv şi negativ. Uneori, respondenții dezvoltă aşa-numitul „set al răspunsului încuviințat". Aceasta înseamnă că respondenții vor tinde să răspundă cu „sunt de acord" la majoritatea itemilor. Or, pentru a evita o asemenea tendință, trebuie să includeți şi itemi care sunt formulați în direcție opusă. De exemplu, într-un test ce măsoară

Copyright © DEPARTAMENT ID 2008

depresia, regăsim şi itemi formulaţi pozitiv, de genul „M-am simţit deprimat" dar şi itemi formulaţi în direcţie opusă, de genul „M-am simţit plin de speranţă privind viitorul".

Teme de reflecţie/autoevaluare:
1. Care credeţi că sunt avantajele şi dezavantajele în cazul în care testul este foarte scurt şi în cazul în care testul este foarte lung.
2. La ce categorii de instrumente psihodiagnostice se aplică cel mai bine ghidul lui DeVellis ?

2. Tipuri de analiză de itemi

Pentru a înţelege mai bine ce înseamnă analiza itemului vom apela la o comparaţie extrem de utilă şi de interesantă oferită de D. Laveault şi J. Gregoire, 1997 (apud Stan, A., 2002). Aceşti autori compară analiza itemilor cu repetiţia unei orchestre în care instrumentele trebuie să cânte armonios şi să intervină într-un mod foarte precis. „Totul trebuie să producă o senzaţie muzicală aparte corespunzând intenţiilor compozitorului şi ale dirijorului".

Analiza itemului apare ca fiind necesară atunci când construim un test nou şi atunci când dorim să facem o nouă revizie a unui test care a fost folosit o anumită perioadă de timp dar care dă semne că unii itemi nu mai sunt valizi. Analiza itemului presupune o analiză *cantitativă,* cât şi una *calitativă.* Cea cantitativă presupune utilizarea unor procedee statistice pentru calcularea unor indici, precum, de exemplu, indexul dificultăţii itemului şi indexul discriminării itemului. Analiza calitativă vizează, pe de o parte, conţinutul şi, pe de altă parte, formatul itemilor.

Criteriile folosite pentru a selecta itemii depind foarte mult de obiectivele urmărite de către constructorul de test. De exemplu (Cohen, J.R., Swerdlik, E.M., 2005) un constructor de teste poate considera ca fiind cei mai buni itemi cei care contribuie în mod optim la fidelitatea internă a testului, în timp ce alt constructor de teste poate dori să proiecteze un nou test cu cea mai mare valoare posibilă a validităţii referitoare la criteriu.

Printre mijloacele pe care le poate folosi pentru analiza itemului pot fi menţionate următoarele: a) un index al dificultăţii itemului; b) un index al discriminării itemului; c) un index al fidelităţii itemului; d) un index al validităţii itemului. Noi ne vom ocupa, în capitolele următoare, mai ales de primele două forme de index.

Analiza calitativă a itemului presupune utilizarea unor variate proceduri nonstatistice (deoarece nu poate fi vorba despre o abordare numerică) de analiză, după cum am mai afirmat şi mai sus, a conţinutului şi formatului. Trebuie găsită cea mai bună formulare a itemului. De aceea trebuie iniţiate cercetări în această direcţie, în sensul că cei care răspund la un test nou sunt rugaţi să răspundă la unele întrebări de genul celor prezentate în tabelul următor (Cohen, J.R., Swerdlik, E.M., 2005):

Copyright © DEPARTAMENT ID 2008

Problema	Exemplu de întrebare
Sensilibilitatea culturală	Ai simțit că unii itemi au fost discriminatori în raport cu anumite grupuri de subiecți. Dacă da, de ce ?
Validitatea de fațadă	Pare testul că măsoară ceea ce aşteptați să măsoare ? Dacă nu, ce este contrar aşteptă-rilor ?
Aplicantul testului	V-a afectat în vreun fel conduita aplicantului testului performanțele dvs. ? Dacă, da, cum ?
Mediul de aplicare	A afectat în vreun fel condițiile din sala de testare performanța dvs. la test ? Dacă da, cum ?
Corectitudinea testului	Credeți că testul este corect în raport cu ceea este considerat că măsoară ? De ce sau de ce nu ?
Limbajul testului	În raport cu care dintre instrucțiuni sau alte aspecte scrise ale testului ați avut dificultăți de înțelegere ?
Lungimea testului	Ce ați simțit, referitor la lungimea testului, în privința: a) timpul necesar pentru completare; b) numărul de itemi ?
Ghicitul răspunsului	Ați ghicit la vreunul dintre itemii testului ? Care credeți că este procentul itemilor la care ați ghicit ? Ați utilizat o strategie particulară pentru a ghici sau ați ghicit la întâmplare ?
Integritatea subiectului	Credeți că s-a putut trişa la acest test ? Dacă da, descrieți metodele ce credeți că au fost utilizate.
Starea mentală şi fizică a subiectului	Cum ați descrie starea dvs. mentală la începu-tul testării ? Credeți că această stare a afectat în vreun fel rezultatele obținute ? Dacă da, cum ? Cum descrieți starea dvs. fizică la înce-putul testării ? Credeți că această stare a afec-tat în vreun fel rezultatele obținute ? Dacă da, cum ?
Starea mentală şi fizică a sbiectului	Cum ați descrie starea dvs. mentală pe parcur-sul testării ? Credeți că această stare a afectat în vreun fel rezultatele obținute ? Dacă da, cum ? Cum ați descrie starea dvs. fizică pe parcursul testării ? Credeți că această stare a afectat în vreun fel rezultatele obținute ? Dacă da, cum ?
Impresia generală a subiectului	Care este impresia dvs. generală privind acest test ? Ce sugestii ați oferi constructorului de tes pentru îmbunătățire ?

Copyright © DEPARTAMENT ID 2008

Preferinţele subiectului	Aţi găsit vreuna din părţile testului ca având un efect educaţional, distractiv sau în alt fel recompensator ? Ce v-a plăcut şi ce nu v-a plăcut în mod special la acest test ? Aţi găsit vreuna din părţile testului care să provoace anxietate, condescendenţă sau să supere în alt fel ? De ce ?
Pregătirea subiectului	Cum v-aţi pregătit pentru acest test ? Dacă ar fi să sfătuiţi pe alţii cum să se pregătească pentru test, ce i-aţi spune ?

Întrebările pot fi prezentate fie în oral, fie în scris. În plus, în funcţie de obiectivele urmărite de utilizatorul de test, întrebările pot fi prevăzute cu mai multe modalităţi de răspuns, cum ar fi, răspuns dihotomic, de tipul „Adevărat-fals" sau cu mai multe variante de răspuns. De asemenea, pentru mai multă siguranţă este recomandabil ca să se asigure confidenţialitatea răspunsurilor subiecţilor.

Teme de reflecţie/autoevaluare

1. De ce este necesară analiza cantitativă a itemului?
2. De ce este necesară analiza cantitativă a itemului ?

Întrebări de autoevaluare:

1. Ce este, de fapt, itemul unui test ?
2. Ce forme alternative există pentru itemi şi analiza de itemi ?
3. Ce efecte poate avea scurtarea lungimii unui test ?
4. Ce efecte poate avea mărirea dimensiunii testului ?
5. Câţi itemi poate avea un test?
6. De cine depinde numărul itemilor ?
7. Cum se poate construi un nou test ?
8. În ce constă analiza cantitativă a itemilor ?
9. În ce constă analiza calitativă a itemilor ?
10. Ce tipuri de index pot fi calculate ?
11. Cum se poate asigura feed-back-ul pentru constructorul de test referitor la aspectele calitative ale noului test ?
12. De ce este bine ca, atunci când verificăm aspectele calitative ale unui test nou, să păstrăm confidenţialitatea răspunsurilor ?

BIBLIOGRAFIE:

1. COHEN, J.R., Swerdlik, E.M. (2005). *Psychological Tesing and Assessment.An Introduction to Tests and Measurement*. McGraw-Hill International Edition.
2. CLINCIU, A. I. (2005). *Psihodiagnostic*. Braşov: Ed. Universităţii „Tansilvania".
3. FRIEDENBERG, L. (1995). *Psychological testing. Design, Analysis, and use*. Allyn & Bacon.
4. KAPLAN, M.R., SACCUZZO, P.D. (2005). *Psychological Testing. Principles, Applications, and Issues*. Thomson Wadsworth.
5. MINULESCU, M. (2003). *Teorie şi practică în psihodiagnoză*. *Testarea intelectului*. Bucureşti: Editura Fundaţiei România de Mâine.

Copyright © DEPARTAMENT ID 2008

6. STAN, A. (2002). *Testul psihologic.Evoluţie, construcţie, aplicaţii*. Iaşi: Polirom.

Copyright © DEPARTAMENT ID 2008

Unitatea de învățare nr. 2

Gradul de dificultate al itemului

Cuprins:

1. La ce se referă gradul de dificultate al itemului ?

2. Cum se calculează dificultatea itemului

Obiective:
La sfârşitul acestei unități de învățare studenții vor fi capabili să:

- explice în ce constă gradul de dificultate al unui item;
- cunoască ce se întâmplă în cazul în care itemii sunt foarte uşori sau foarte dificili;
- înțeleagă de ce valoarea cea mai bună a nivelului de dificultate este p = 0,50;
- poată calcula nivelul de dificultate al unui item şi, respectiv, al unui test

1. La ce se referă gradul de dificultate al unui item ?

Trebuie să recunoaştem că întrebarea nu este prea dificilă. Un răspuns îl putem oferi imediat, adică se pune problema cât de greu sau cât de uşor este itemul pentru subiecți. Da, dar subiecții sunt destul de diferiți în ceea ce priveşte însuşirea sau trăsătura psihică pe care vrea să o măsoare testul. Şi testul tocmai asta urmăreşte şi anume, să scoată în evidență diferențele interindividuale. Un item poate fi dificil datorită faptului că niciunul dintre subiecții pe care se aplică nu îl poate rezolva. Iată, de exemplu, dacă ne-am adresa cu un test unui grup de 100 de studenți la psihologie, din care face parte un item de genul: "Cât fac 456.932 x 297465 ?" În mod sigur nu va putea răspunde nimeni corect, ceea ce înseamnă că itemul este foarte dificil. În acest caz, $p = 0$, p însemnând procentul celor care răspund la test. Dar dacă, în cadrul testului, există şi un item de genul "Cât fac 4 x 2 ?" Evident că, în acest caz, toți studenții vor răspunde correct, deoarece la o asemenea întrebare răspund corect şi elevi din ciclul primar. Aici $p = 100$ şi asta înseamnă că toți subiecții răspund correct. În ambele cazuri, itemii sunt nefolositori, deoarece ei nu ne oferă niciun fel de informație referitoare la deprinderile de calcul ale subiecților.

Să presupunem acum că, în caz de aplicare a testului, avem un item la care răspund corect doar 10 subiecți (caz în care $p = 10$). Acest item prezintă un nivel înalt de dificultate. Şi în cazul în care ar răspunde 90 de subiecți, itemul ar avea un nivel redus de dificultate. Dacă avem în vedere mijlocul scalei, respectiv, $p = 50$, cu cât valoarea lui p va creşte, cu atât nivelul de dificultate al itemului va scădea şi, invers, cu cât valoarea lui p va fi mai mică, cu

Copyright © DEPARTAMENT ID 2008

atât nivelul de dificultate al itemului va creşte. Deci, ar rezulta, chiar şi la o analiză mai superficială, că valoarea optimă a nivelului de dificultate al unui item este 0.50.

Dar pot fi aduse dovezi mai serioase. Astfel, dacă la un test aplicat pe 100 de subiecți, reuşesc 50 dintre ei, înseamnă că, firesc, ceilalți 50 nu reuşesc. În felul acesta, putem să comparaăm pe fiecare subiect din grupul de 50, care au reuşit la item, cu fiecare dintre cei 50 de subiecți care nu au reuşit la item. Avem, astfel, 50 x 50 sau 2500 comparații perechi sau *biți ai informației diferențiale* (Anastasi, A., 1988). Oricare altă combinație nu conduce la o asemenea valoare. Pentru mai multă siguranță verificați tabelul de mai jos:

0	**x**	**100**	**=**	**0**
10	x	90	=	900
20	x	80	=	1600
30	x	70	=	2100
40	x	60	=	2400
50	**x**	**50**	**=**	**2500**
60	x	40	=	2400
70	x	30	=	2100
80	x	20	=	1600
90	x	10	=	900
100	**x**	**0**	**=**	**0**

Temă de reflecție/autoevaluare

Alcătuiți un test care să cuprindă mai mulți itemi, fiecare dintre ei având un nivel de dificultate diferit, dar, pe ansamblu, nivelul de dificultate al testului să fie 0,50.

2. Cum se calculează dificultatea itemului.

Analizând tabelul de mai sus, am putea trage o concluzie foarte simplă şi anume, testul cel mai bun este cel care este trecut de 50 % dintre subiecți, adică fiecare dintre itemii săi are nivelul de dificultate, $p = 50$. A. Anastasi recomandă, însă, ca să fie selectați itemii cu nivele de dificultate diferite dar a căror medie a dificultății să fie 0,50. Totodată, această autoare vorbeşte de utilitatea unor *scale de interval* pentru determinarea nivelului de dificultate al unui item (Anastasi, A., 1988, 1996).

Dacă presupunem o distribuție normală a unei trăsături psihice măsurată de un anumit item, nivelul de dificultate al itemului poate fi exprimat în funcție de o scală cu unități de interval egale prin referirea la o tablă a frecvențelor curbei normale. Ştim că aproximativ 34% din cazuri într-o distribuție normală cad între medie şi distanța de 1σ în oricare direcție. Dacă, de exemplu, un item a fost trecut de 84% dintre subiecți, înseamnă că acest procent cuprinde 50% din partea superioară a distribuției şi 34% din partea inferioară a acesteia. Deci, în concluzie, acest item cade în 1σ sub medie.

Copyright © DEPARTAMENT ID 2008

Un item trecut de 16% din cazuri va cădea în 1σ deasupra mediei, pentru că, deasupra acestui punct, există 16% sin cazuri (50 – 34 = 16). Iar un item la care au reuşit 50% din cazuri, cade exact pe medie şi ar avea, astfel, valoarea *0* pe această scală.

Deoarece dificultatea itemului exprimată în funcție de curba normală, ce presupune deviația standard ca distanță, implică valori negative şi zecimale, ele sunt convertite într-o scală uşor de mânuit şi care este desemnată prin litera grecească delta Δ. Relația dintre Δ şi curba normală – distanțele σ (valori z) este următoarea:

$$\Delta = 13 + 4z$$

În această formulă 13 şi 4 sunt constante alese pentru a oferi o scală care să elimine valorile negative şi zecimalele.

Un item trecut de aproape 100% din subiecți (99,87%), căzând în - 3σ, ar avea o valoare a lui Δ = 1 (aşa cum reiese din formulă: 13 + (4)(-3) = 1)

La cealaltă extremă, un item care a fost trecut de mai puțin de 1% din subiecți (0,13%), ar cădea în + 3σ şi ar avea valoarea lui Δ = 25 (aşa cum apare din formulă: 13 + (4) (3) = 25). Şi, în sfârşit, un item ce ar cădea pe medie va avea valoarea lui Δ = 0 (aşa cum reiese din formulă:

13 + (4) (0) = 25.

Se poate trage concluzia că Δ este o scală în care, practic, toți itemii cad, în ceea ce priveşte nivelul de dificultate, între valorile 1 şi 25, cu o medie = 13.

Referitor la calcularea nivelului de dificultate al unui item mai este o problemă extrem de importantă. La un item se poate răspunde corect şi din întâmplare sau, cum se mai spune mai simplu, ghicind răspunsul. De exemplu, dacă la un item se răspunde dihotomic, prin "Da" sau "Nu", "Adevărat" sau "Fals", subiectul poate ghici 50 % din răspunsurile corecte. Sau, dacă la un item, există 4 variante de răspuns, printre care şi cel correct, subiectul poate ghici răspunsul corect în procent de 25%.

Nivelul de dificultate optim al itemilor se află la jumătatea diferenței dintre numărul maxim de subiecți ce răspund correct (100%) şi nivelul reuşitei bazată numai pe şansă (noroc). Astfel, nivelul optim de dificultate al unui item prevăzut cu patru variante de răspuns, printre care şi cel corect este de aproximativ 0,625: Pentru a se ajunge la această valoare, se parcurg următorii paşi (Kaplan, M.R., Saccuzzo, P.D., 2005):

1. Găsiți jumătate din diferența dintre 100% succes şi performanța pe bază de şansă

$$\frac{100 - 0{,}25}{2} = \frac{0{,}75}{2} = 0{,}375$$

2. Adăugați această valoare probabilității de a obține răspunsul corect pe bază de şansă

$$0{,}375 + 0{,}25 = 0{,}625$$

O metodă mai simplă pentru obținerea aceluiaşi rezultat este de a adăuga 1,00 la probabilitatea de a reuşi pe bază de şansă şi apoi de a divide la 2. Astfel:

$$\frac{1{,}00 + 0{,}25}{} = 0{,}625$$

Temă de reflecție/autoevaluare:

Copyright © DEPARTAMENT ID 2008

Pentru un item există două variante de răspuns. Având în vedere probabilitatea de a raspunde corect pe bază de şansă, care ar trebui să fie nivelul de dificultate al acestui item ?	2

Întrebări de autoevaluare:

1. Ce înseamnă că un item este prea dificil ?
2. Ce i se poate imputa unui asemenea item ?
3. Ce înseamnă că un item este prea uşor ?
4. Ce i se poate imputa unui asemenea item ?
5. Cum se poate pune în evidenţă nivelul de dificultate al unui item ?
6. Care este valoarea optimă a nivelului de dificultate al unui item ?
7. La ce se referă sintagma „biţi ai informaţiei diferenţiale ?”
8. Care este utilitatea unei scale de interval ?
9. Care sunt principalele valori ale unei scale de interval ?
10. Care sunt paşii ce trebuie parcurşi pentru a calcula exact nivelul de dificultate al unui item ?

BIBLIOGRAFIE:

1. ALBU, M. (1998). *Construirea şi utilizarea testelor psihologice*. Cluj: Editura Clusium.
2. ALBU, M., PITARIU, H. (1993). *Proiectarea testelor de cunoştinţe şi examenul asistat de calculator*. Cluj-Napoca: Editura Casa cărţii de ştiinţă.
3. COHEN, J.R., SWERDLIK, E.M. (2005). *Psychological Testing and Assessment. An Introduction to Tests and Measurement*. McGraw-Hill International Edition.
4. CLINCIU, A.I. (2005). *Psihodiagnostic*. Braşov: Ed. Universităţii „Transilvania”
5. FRIEDENBERG, L. (1995). *Psychological Testing. Design, Analzsis, and use.* Allyn & Bacon.
6. KAPLAN, M.R., SACCUZZO, P.D. (2005). *Psychological Testing. Principles, Applications, and Issues*. Thomson Wadsworth.
7. MINULESCU, M. (2003). *Teorie şi practică în psihodiagnoză*. *Testarea intelectului*. Bucureşti: Editura Fundaţiei România de Mâine.
8. STAN, A. (2002). *Testul psihologic.Evoluţie, consatrucţie, aplicaţii*. Iaşi: Polirom.

Copyright © DEPARTAMENT ID 2008

Unitatea de învățare nr. 3
Capacitatea de discriminare a itemului
Cuprins:

1. Ce este analiza discriminării itemului ?
2. Cum se calculează indexul de discriminare al itemului

Obiective:
La sfârşitul acestei unități de învățare, studenții vor fi capabili să:

- cunoască în ce constă capacitatea de discriminare a itemului;
- înțeleagă ce legături există cu nivelul de dificultate al itemului;
- cunoască modalitățile de calcul a indexului de discriminare a itemului
- poată exemplifica uşor

1. Ce este analiza discriminării itemului ?

Dacă nivelul de dificultate al unui item arată cât de uşor sau cât de dificil este un item, nivelul de discriminare arată în ce măsură subiecții, care obțin rezultate bune la întregul test, reuşesc şi la acest item şi, pe de altă parte, în ce măsură subiecții, care obțin rezultate slabe la întregul test, nu reuşesc la acest item. După cum deja ştiți, fiecare test măsoară o anumită însuşire sau caracteristică psihică şi fiecare item din structura sa trebuie să măsoare acelaşi lucru. Deci, la subiecții care obțin scoruri mari la întregul test, înseamnă că la ei nivelul de dezvoltare şi de funcționalitate al acelei însuşiri sau capacități psihice este foarte înalt. Iar la subiecții care obțin scoruri mici la întregul test, acest nivel este scăzut. Această diferențiere o vor realiza şi itemii, deoarece fiecare dintre ei măsoară acelaşi lucru precum testul luat în întregul său. Deci, pentru cei care au un nivel ridicat de dezvoltare şi de funcționalitate al unei însuşiri sau capacități psihice scorurile la întregul test vor fi mari şi ei vor reuşi, în general, şi la itemii testului. Invers, la subiecții la care nivelul de dezvoltare şi de funcționalitate al aceleiaşi însuşiri sau capacități psihice supuse măsurării este redus, vom regări scoruri slabe la întregul test şi eşecuri la itemii ce îl alcătuiesc. Şi, dacă tot nu s-a înțeles, să apelăm la o comparație, care, desigur, este contraindicată. Cântărim o persoană cu un cântar mare şi vedem că are 100 de kg. Apoi îl cântărim cu mai multe cântare mici. Vom constata că persoana respectivă va avea tot 100 de kg. Rugăm o altă persoană să accepte să fie cântărită şi constatăm că va avea 40 de kg. Apoi, folosind mai multe cântare mici, o să vedem că persoana respectivă va avea tot 40 de kg. Deci şi cântarul mare şi cântarele mici au capacitatea de a discrimina persoanele în funcție de greutate.

Revenind la test, în cazul în care un test este foarte bine construit, subiectul bun, dotat, va avea un scor mare la test şi va reuşi la toți itemii iar un subiect slab, nedotat, va avea un scor foarte slab la test şi nu va reuşi aproape la toți itemii. Testul, însă, aşa după cum am spus de atâtea ori, prezintă anumite imperfecțiuni, care, se răsfrâng, desigur, şi asupra unor itemi. De aceea, practic, nu întâlnim situația în care toți subiecții buni să răspundă corect la toți itemii şi subiecții slabi să nu răspundă corect la niciun item. Pe de altă

Copyright © DEPARTAMENT ID 2008

parte, nici subiecţii buni şi nici subiecţii slabi nu au exact acelaşi nivel de dezvoltare al însuşirii sau capacităţii psihice supuse măsurării.

De aceea, testul trebuie să evidenţieze cât mai exact posibil acest nivel de dezvoltare. Şi atunci, subiecţii, deşi se află într-un anumit registru de dezvoltare al însuşirii sau capacităţii psihice respective (înalt sau scăzut), vor trebui să fie şi ei diferenţiaţi sau, ca să fim consecvenţi cu terminologia, discriminaţi. Aceasta înseamnă că cei buni, dar nu excelenţi, vor mai greşi la unii itemi şi, pe de altă parte, cei slabi, dar nu extrem de slabi, vor reuşi la unii itemi. Pot să apară şi anumite paradoxuri. Cei foarte buni să nu reuşească la un anumit item şi cei slabi să reuşească la acesta. Putem înţelege uşor că, în asemenea cazuri, este ceva în neregulă cu itemii respectivi, ceea ce ne obligă la o reevaluare a acestora. Ei nu pot să rămână aşa cum sunt în structura testului, deoarece s-ar putea ca ei să măsoare ceva ce nu are legătură cu testul din care fac parte.

Temă de reflecţie/autoevaluare:
Nivelul de dificultate optim al unui item este 0,50. Ce putem spune în legătură cu capacitatea lui de discriminare?

2. Indexul de discriminare al itemului

În practică există foarte multe proceduri ce pot fi folosite pentru calcularea indexului de discriminare al itemului. Cea mai frecvent utilizată este cea care ia în atenţie 3 grupe de subiecţi, care, la rândul lor, alcătuiesc numărul total al subiecţilor ce răspund la un test (N): a) grupul celor care au obţinut răspunsuri bune la test; b) grupul celor care au obţinut răspunsuri mediocre la test; c) grupul celor care au obţinut rezultate slabe la test. Primul grup poate fi notat cu U (de la Upper), al doilea, cu M (de la Middle) şi, al treilea, cu L (de la Lower). După alţi autori, însă, este suficient să luăm în atenţie numai grupul subiecţilor care au răspuns mai bine la test (U) şi grupul subiecţilor care au obţinut rezultate slabe la test (L). În acest caz se vorbeşte de *metoda grupelor extreme* (Friedenberg, L., 1995) şi ea este utilizată cel mai frecvent pe testele de abilităţi, testele de personalitate, testele de interese şi testele de atitudini.

Să luăm, mai întâi, un exemplu oferit de A. Anastasi (1988, 1996). În urma aplicării unui test unui grup de 60 de studenţi, îi împărţim, în funcţie de rezultate, în cele 3 grupuri, fiecare cuprinzând câte 20 de studenţi, respectiv, 33%. (Trebuie să menţionăm, în această paranteză că, după alţi autori, procentul subiecţilor ce aparţin grupelor U şi L poate fi 25% sau 27%. În această privinţă, în 1979, Allen, M.J. şi Yen W.M., apud, Cohen, J.R., Swerdlik, E.M., 2005, ne asigură că, pentru cele mai multe aplicaţii, orice procentaj între 25 şi 27 va produce estimări similare). Apoi înregistrăm răspunsurile corecte pentru fiecare item oferite de către studenţii ce aparţin celor trei grupe. Se alcătuieşte un tabel în felul următor:

Item	U	M	L	Dificultate U+M+L	Discriminare U - L
1	15	9	7	31	8
2	20	20	16	56*	4

Copyright © DEPARTAMENT ID 2008

3	19	18	9	46	10
4	10	11	16	37	- 6*
5	11	13	11	35	0*
6	16	14	9	39	7
7	5	0	0	5*	5

În acest tabel sunt prezentate rezultatele pentru numai 7 itemi din cadrul testului. După cum se poate observa, nivelul de dificultate al itemului se află însumând numărul subiecţilor din cele 3 grupe care răspund corect la toţi itemii. Pe de altă parte, valoarea discriminativă a fiecărui item poate fi aflată făcând diferenţa dintre numărul subiecţilor care răspund corect în grupul U şi numărul subiecţilor care răspund corect în grupul L. Itemii ce prezintă probleme sunt cei care sunt prevăzuţi cu steluţă. Astfel, itemii 2 şi 7 fac notă discordantă faţă de ceilalţi itemi, în sensul că itemul 2 este prea uşor (56 de subiecţi răspund corect) iar itemul 7 este prea greu (numai 5 subiecţi răspund corect).

În ultima coloană itemii 4 şi 5 prezintă probleme, în sensul că, deşi satisfăcători în ceea ce priveşte nivelul de dificultate, itemul 4 prezintă o valoare discriminativă negativă iar itemul 5 prezintă o valoare discriminativă egală cu zero. Aceşti itemi nu pot să rămână aşa în structura testului existent sau nou construit, deoarece ei, practic, nu prezintă niciun fel de utilitate. Este clar că trebuie să fie supuşi unor serioase acţiuni de analiză şi evaluare sau, pur şi simplu, pot fi scoşi definitiv din componenţa testului.

A. Anastasi arată că, în cazul în care numărul subiecţilor care reuşesc la fiecare item din grupele U şi L este exprimat în procente, indexul discriminării poate fi calculat prin diferenţa dintre cele două procentaje, precum în tabelul următor:

Item	Procent reuşită U grup	Procent reuşită L grup	Indicele discriminării U - L
1	75	35	40
2	100	80	20
3	95	45	50
4	50	80	- 30*
5	55	55	0*
6	80	45	35
7	25	0	25

După cum se poate observa din tabel, aceeaşi itemi prezintă probleme, respectiv itemii 4 şi 5. Aceştia trebuie să fie renalizaţi pentru a se vedea dacă mai rămân în structura testului. Deci şi în cazul procentajului diferenţa dintre U şi L constituie valoarea indexului discriminării pentru fiecare item. Acesta poate fi evidenţiat în mai multe forme, precum: U-L, ULI, ULD, D sau d.

Analizând datele din ambele tabele de mai sus, constatăm că, în afară de itemii care ridică probleme privind valoarea lor discriminativă, pentru ceilalţi itemi valorile sunt diferite, de aceea apare întrebarea care dintre ele este cea mai bună ? Referitor la metoda utilizării procentajelor, R.L. Ebel (apud Stan, A., 2002) a propus, în 1965, o scală de repere pentru interpretarea valorică a indicelui de discriminare D:

Copyright © DEPARTAMENT ID 2008

Interval valoric al lui D	Interpretarea discriminării
0,40 şi mai mult	discriminare foarte bună
de la 0,30 la 0,39	discriminare bună
de la 0,20 la 0,29	discriminare slabă
de la 0,10 la 0,19	discriminare de limită
sub 0,10	fără utilitate

Atunci, însă, când se foloseşte metoda efectuării diferenţei dintre numărul răspunsurilor corecte oferite de subiecţii ce aparţin celor două grupe, unii autori (Anastasi, A., 1988, 1996) fac următoarele precizări: a) itemii cu valoare discriminativă scăzută sunt cei care au o valoare a diferenţei de la 3 în jos (diferenţa dintre U şi L) şi asta când grupurile comparate au aceeaşi dimensiune; b) cu cât grupurile sunt mai mari, ne aşteptăm ca să apară diferenţe mai mari, datorate şansei legate de un item nediscriminativ.

Pentru a înţelege şi mai bine utilitatea calculării valorii indexului discriminării itemului vom lua şi un alt exemplu (Cohen, J.R., Sewrdlik, E.M., 2005). A fost aplicat un test de cunoştinţe pe un număr de 119 studenţi. În urma rezultatelor obţinute au fost alcătuite cele două grupe, U şi L, fiecare dintre ele cuprinzând 27%, respectiv, 32 de studenţi. O secvenţă vizând primii 5 itemi este prezentată în tabelul de mai jos.

Item	U	L	U – L	n	d = U - L/n
1	20	16	4	32	0,13
2	30	10	20	32	0,63
3	32	0	32	32	1,00
4	20	20	0	32	0,00
5	0	32	-32	32	- 1.00

În primul rând trebuie menţionat faptul că valoarea indicelui discriminprii testului se întinde între – 1,00 şi + 1,00. Valoarea de – 1,00 indică faptul că toţi subiecţii ce aparţin grupului U nu reuşesc la test şi toţi subiecţii care aparţin grupului L reuşesc la test. Este o situaţie paradoxală ce solicită de urgenţă revizia sau eliminarea itemului din cadrul testului. Valoarea de + 1,00 arată că toţi subiecţii ce aparţin grupului U reuşesc la test şi toţi subiecţii care aparţin grupului U nu reuşesc la test. Când o aceeaşi proporţie a reuşitei o regăsim la ambele grupe (U şi L) valoarea discriminativă a testului este egală cu zero. În acest caz itemul nu discriminează deloc între subiecţi, de aceea este inutil.

Din aceste exemple reiese clar faptul că există o strânsă legătură între nivelul de dificultate al itemului şi capacitatea lui de discriminare. A. Anastasi (1988, 1996), ne oferă un tabel care evidenţiază relaţia dintre nivelul de dificultate al itemului şi valoarea maximă a indexului discriminării.

Procentul subiecţilor ce reuşesc la item	Valoarea maximă a lui D

Copyright © DEPARTAMENT ID 2008

100	0
90	20
80	40
70	60
60	80
50	100
40	80
30	60
20	40
10	20
0	0

Cum de apar asemenea valori ale lui D ? Când toţi subiecţii din grupul U reuşesc la item (50/50 = 100%) şi toţi subiecţii din grupul L reuşesc la acest item (0/50 = 0), atunci valoarea lui D (U – L) = 0. Când niciun subiect din grupul U (0/50 = 0) şi niciun subiect din grupul L (0/50 = L) nu reuşeşte la un item, atunci valoarea lui D este tot 0. Când toţi subiecţii din grupul U (50/50 = 100%) şi niciun subiect din grupul L (0/50 = 0 %) atunci valoarea lui D = 100. Şi aceasta este cea mai mare valoare a lui D. Dacă reuşesc la item 90 % din subiecţi, asta înseamnă că au reuşit toţi din grupul U (50/50 = 100%) şi 40 din grupul L (40/50 = 80%). În consecinţă, valoarea lui D (U – L) va fi 20. Dacă reuşesc la item 80% din subiecţi, asta înseamnă că toţi subiecţii din grupul U (50/50 = 100%) şi numai 30 (30/50 = 60) din grupul L reuşesc la item. Făcând diferenţa vom constata că valoarea lui D = 40. Ş.a.m.d.

Teme de reflecţie/autoevaluare:

1. Care este explicaţia situaţiei în care valoarea lui D (indicele de discriminare al itemului) este negativă ?
2. De ce itemul al cărui nivel de dificultate este 0,50 are cea mai mare valoare a indicelui de discriminare ?

Întrebări de autoevaluare:

1. Ce este analiza discriminării itemului ?
2. Ce proceduri de calculare a indexului discriminării cunoaşteţi ?
3. Care este procedura recomandată de A. Anastasi ?
4. În ce constă metoda grupurilor extreme ?
5. Care este procentul subiecţilor din grupurile U şi L recomandat de autori ?
6. Ce trebuie făcut cu itemii ai căror valoare discriminativă este foarte scăzută ?
7. Care sunt simbolurile recomandate pentru indexul discriminării itemului ?
8. Care este opinia lui R.L. Ebel privind intervalul valoric al lui D ?
9. Dar a lui A. Anastasi ?
10. Explicaţi ce se întâmplă atunci când valoarea lui D este – 1,00 şi + 1,00

BIBLIOGRAFIE:

Copyright © DEPARTAMENT ID 2008

1. ALBU, M. (1998). *Construirea şi utilizarea testelor psihologice*. Cluj: Editura Clusium.
2. ALBU, M., PITARIU, H. (1993). *Proiectarea testelor de cunoştinţe şi examenul asistat de calculator*. Cluj-Napoca: Editura Casa cărţii de ştiinţă.
3. COHEN, J.R., SWERDLIK, E.M. (2005). *Psychological Testing and Assessment. An Introduction to Tests and Measurement*. McGraw-Hill International Edition.
4. CLINCIU, A.I. (2005). *Psihodiagnostic*. Braşov: Ed. Universităţii „Transilvania"
5. FRIEDENBERG, L. (1995). *Psychological Testing. Design, Analysis, and use.* Allyn & Bacon.
6. KAPLAN, M.R., SACCUZZO, P.D. (2005). *Psychological Testing. Principles, Applications, and Issues*. Thomson Wadsworth.
7. MINULESCU, M. (2003). *Teorie şi practică în psihodiagnoză*. *Testarea intelectului*. Bucureşti: Editura Fundaţiei România de Mâine.
9. STAN, A. (2002). *Testul psihologic*.*Evoluţie, consatrucţie, aplicaţii*. Iaşi: Polirom.

Copyright © DEPARTAMENT ID 2008

MODULUL V

ETALONAREA ŞI STANDARDIZAREA TESTELOR

Unitatea de învățare nr. 1

Etapele etalonării şi standardizării testelor

Cuprins:

1. Definirea conceptelor şi etapele construirii testelor
2. Modele de scale utilizate în etalonare şi standardizare

Obiective:

La sfârşitul acestei unități de învățare studenții vor fi capabili să:

- definească corect concepte precum: construcția testului, etalonare, standardizare;
- precizeze care sunt etapele parcurse în vederea etalonării unor noi teste;
- diferențeze între acțiunea de etalonare şi cea de re-etalonare a testelor;
- înțeleagă bine ce înseamnă reprezentaivitatea eşantionului.

1. Definirea conceptelor

De ce trebuie să fie testele etalonate şi standardizate ? Să presupunem că vom construi un nou test de inteligență pentru studenții din România şi numai un singur student nu îl includem în lot. Toate rezultatele obținute le vom grupa în mai multe clase sau categorii, plecând de la cele mai bune şi ajungând la cele mai slabe. Desigur, vor exista mulți studenți care vor obține acelaşi rezultat, de aceea trebuie să luăm în considerație şi frecvența răspunsurilor. Deci se face un fel de clasament, pe primele locuri fiind rezultatele cele mai bune şi, pe ultimile locuri, rezultatele cele mai slabe. Singurului subiect care nu a fost inclus în lot îi aplicăm noul test de inteligență creat şi, apoi, rezultatul pe care îl obține îl raportăm la „clasamentul" rezultatelor obținute de toți subiecții incluşi în lot. Vom vedea dacă studentul respectiv se plasează pe o poziție mai spre vârful clasamentului, mai spre mijloc sau mai spre baza acestuia. Deci comparăm rezultatul subiectului cu rezultatele întregului lot.

Pornind de la acest exemplu, trebuie să deosebim între situația în care un test deja există şi cea în care se construieşte un test nou. Pentru testul deja existent nu se pune problema etalonării lui, presupunând că există un etalon, ci, eventual, a reetalonării lui, în cazul în care etalonul nu mai corespunde. De asemenea, dacă testul provine din altă țară, deşi el are un etalon (în limba engleză se foloseşte termenul de „norme" – norms - în loc de etalon), trebuie să fie re-etalonat, trecând, mai întâi, prin faza traducerii şi adaptării. De aici putem trage concluzia conform căreia construirea unui test nou presupune automat şi etalonarea lui însă etalonarea nu presupune automat şi construirea testului, deoarece acesta poate exista deja (construit de alții) sau este luat din altă parte.

Copyright © DEPARTAMENT ID 2008

O altă noțiune care trebuie să o luăm în atenție este cea de *revizie* a testelor. Mai întâi, să dăm exemple. Bateria Stanford-Binet, care actualmente, este folosită sub denumirea de Stanford-Binet Intelligence Scale (fifth edition) a trecut prin următoarele etape de revizuire: 1905 – Scala de inteligență a lui A. Binet şi T. Simon; 1908 – prima revizie a Scalei de inteligență Binet-Simon; 1911 – a doua revizie a Scalei de inteligență Binet-Simon; 1916 – apare versiunea Stanford-Binet Intelligence Scale, sub îngrijirea lui L. Terman; 1937 – prima revizie a variantei lui Terman; 1960 – a doua revizie; 1972 - a treia revizie; 1985 – a patra revizie (sub îndrumarea lui R.L. Thorndike, E. Hagen şi J. Sattler); 2003 – a cincea revizie, sub îndrumarea lui G. Roid. Bateria WAIS – III, respectiv, Scala de inteligență a lui Wechsler pentru adulți a parcurs, de asemenea, mai multe etape de revizie. Prima variantă a apărut în anul 1939, sub denumirea Wechsler-Bellevue Intelligence Scale I. A doua a apărut în 1942, sub denumirea de The Army Wechsler iar a treia, în 1946, sub denumirea Wechsler Bellvue Intelligence Scale II. În 1955 apare, de fapt, prima variantă a Scalei Wechsler de inteligență pentru adulți (WAIS). Această variantă este revizuită în 1981, devenind, WAIS - R şi, apoi, din nou, în 1997, variantă care există şi în prezent, cunoscută sub denumirea WAIS – III (Scala de inteligență Wechsler pentru adulți – III).

Revizia este o acțiune foarte amplă, care poate dura mult timp (de exemplu, revizia Bateriei Stanford-Binet, fourth edition, a durat între 6 şi 8 ani) şi ea presupune re-analizarea şi re-evaluarea testului din mai multe puncte de vedere: conținut, norme, direcții de administrare şi cotare etc. Părțile testului care sunt mai sensibile la schimbare sunt cele ce presupun folosirea limbajului. Un test preluat din America, de exemplu, trebuie tradus şi adaptat, deoarece acesta poate cuprinde informații care nu sunt specifice şi nu sunt cunoscute de către populația românească. Astfel, sunt unele aspecte legate de istorie, de literatură, de obiceiuri, de tradiții, de preocupări etc. De exemplu, fotbalul american este un sport foarte dur, foarte agresiv față de fotbalul european. Deci, răspunsul pozitiv la întrebarea „Îți place fotbalul ?" înseamnă cu totul altceva pentru un subiect român decât pentru un subiect american. Sau dacă unii itemi presupun operarea cu sistemul de măsurare american, pentru subiecții români trebuie neapărat să fie transpus în sistemul de măsurare folosit în România.

La întrebarea „Când se impune revizuirea unui test ?" putem apela la sugestiile făcute de Asociația Psihologilor Americani – APA – (1996, Standard 3.18): un test poate fi menținut în forma actuală atâta timp cât este folositor şi el trebuie să fie revizuit „atunci când apar schimbări semnificative în domeniul reprezentat sau când condițiile noi pentru utilizare şi interpretare fac testul să fie nepotrivit pentru ceea ce intenționează să măsoare". În opinia altor autori însă, testele trebuie să fie revizuite atunci când există oricare din următoarele condiții (Cohen, R.J., Swerdlik, E.M., 2005):

1. Materialele stimul par a fi demodate şi subiecții nu se pot raporta la ele;
2. Conținutul verbal al testului, incluzând instrucțiunile pentru administrare şi itemii testului, conțin un vocabular depăşit, care nu este imediat înțeles de către subiecți;
3. Întrucât cultura populară se schimbă şi cuvintele capătă noi înțelesuri, anumite cuvinte sau expresii din conținutul itemilor sau din direcțiile de administrare pot fi percepute ca nepotrivite sau chiar ofensatorii pentru un anumit grup particular, de aceea ele trebuie să fie schimbate;
4. Normele testului nu mai sunt adecvate datorită schimbărilor în populația subiecților potențiali;

Copyright © DEPARTAMENT ID 2008

5. Normele testului nu mai sunt adecvate datorită schimbărilor pe linia abilităţilor măsurate şi datorită extinderii lor (în sus sau în jos) dacă acest lucru devine necesar;
6. Fidelitatea sau validitatea testului, precum şi eficacitatea itemilor individuali, pot fi semnificativ îmbunătăţite printr-o revizie.
7. Teoria ce a stat la baza testului original a fost îmbunătăţită semnificativ şi schimbările trebuie să fie reflectate în design-ul şi conţinutul testului.

Privitor la elaborarea şi construirea unui test nou, trebuie să precizăm faptul că este obligatoriu să fie parcurse mai multe etape:

1. Analiza cât mai corectă, cât mai clară a constructului, a însuşirii sau caracteristicii psihice care urmează a fi testată cu instrumentul psihodiagnostic nou. Trebuie definit foarte clar conceptul şi, apoi, operaţionalizat, adică descris din perspectivă comportamentală, actele comportamentale fiind evaluabile şi măsurabile. Trebuie răspuns, cât mai clar, la întrebări de genul: - ce înseamnă a fi sau a nu fi inteligent ? – ce înseamnă a fi sau a nu fi anxios ? – ce înseamnă a avea sa a nu avea memorie logică ? – ce înseamnă a fi sau a nu fi sociabil ?

Am mai vorbit despre asta, dacă vă mai amintiţi, la validitatea privitoare la construct.

2. Imaginarea modalităţilor de simulare a situaţiilor reale în raport cu care subiectul îşi exteriorizează şi îşi obiectivează capacităţile, trăsăturile sale psihice. În această privinţă, un rol foarte important îl are imaginaţia constructorului de test. El trebuie să dea dovadă de inventivitate şi ingeniozitate. Uneori testele impresionează prin forma extrem de originală a itemilor, însă numai aceasta nu asigură automat şi o bună fidelitate şi validitate a testului. De aceea, constructorul de test nu trebuie să acorde mai mare importanţă formei itemilor, în detrimentul caracteristicilor psihometrice ale testelor.

Elaborarea testului presupune, de altfel, respectarea mai multor cerinţe şi anume:

a) alegerea, selectarea materialului utilizat în cadrul testului (cuvinte, propoziţii, numere, figuri, expresii, cuburi, analogii, situaţii problematice, întrebări, evaluări, autoevaluări etc.);
b) formularea instructajului şi a exemplelor pentru familiarizarea subiecţilor cu principiul de construcţie al itemilor testului;
c) precizarea tipului de răspuns (un singur răspuns, mai multe răspunsuri la alegere, sarcină de executat, rezolvarea problemei, exerciţiu logic etc.;
d) elaborarea foii de răspuns, care trebuie să fie folosită numai în formă originală;
e) precizarea modului de cotare (puncte, procente, calificative), precum şi a grilelor ce pot fi folosite;
f) precizarea, dacă este necesar, a timpului cât să dureze testarea, precum şi a faptului dacă este vorba despre un test individual sau test colectiv;

3. Odată construit, testul nou se aplică în forma unei probe oarbe pe un eşantion mai mic, deoarece este o primă verificare a acestuia. Se acordă atenţie unor probleme precum: nivelul de dificultate al itemilor, capacitatea lor de discriminare, dimensiunea testului, dacă este bine formulat şi dacă este suficient instructajul, reacţiile subiectului la test etc.). În funcţie de rezultate, se fac modificările necesare;

4. Etalonarea, etapă ce vizează utilizarea unui ansamblu de procedee statistice pe baza cărora pot fi diferenţiaţi şi clasificaţi subiecţii incluşi în eşantionul folosit în funcţie de rezultatele lor obţinute la un anumit test. În urma unor calcule statistice (trebuie să recunoaştem că, actualmente, acestea se pot realiza uşor folosind calculatorul) rezultatele obţinute de membrii eşantionului sunt grupate în forma unor norme sau etaloane, care, ulterior, sunt folosite drept cadre de referinţă la care raportăm rezultatele unui anumit subiect.

Copyright © DEPARTAMENT ID 2008

În alegerea eşantionului trebuie să respectăm câteva cerinţe obligatorii, mai importante fiind două:

a) reprezentativitatea eşantionului; un eşantion este reprezentativ atunci când el reflectă fidel caracteristicile populaţiei de referinţă. Fiecare membru al eşantionului reprezintă un subgrup din populaţia de referinţă. Procedeele de selectare a eşantionului sunt mai multe, cum ar fi, de exemplu, pasul statistic, alegerea la întâmplare etc. Deci eşantionul este o fracţiune din populaţia de referinţă. De exemplu, dacă noi construim un test de inteligenţă pentru elevii de liceu (populaţia de referinţă), putem să-l aplicăm în vederea elaborării normelor (etalonului) pe un grup mult mai restrâns (eşantionul), care însă trebuie să fie reprezentativ pentru toată populaţia de referinţă. Dacă numărul real al elevilor de liceu este 1.000.000, eşantionul reprezentativ, stabilit în baza pasului statistic (1/1000), ar cuprinde 1000 de elevi. Este ca şi cum am realiza o coloană de 1.000.000 de elevi şi apoi îi selectăm pe cei de pe poziţiile 1, 1001, 2001, 3001, 4001 ş.a.m.d. Fiecare membru al eşantionului reprezintă câte 1000 de elevi din populaţia de referinţă;

b) eşantionul trebuie să fie eterogen în ceea ce priveşte însuşirea sau caracteristica psihică măsurată şi, pe de altă parte, să fie cât mai omogen posibil referitor la alte variabile, în afara celei testate, cum ar fi: vârsta, sexul, pregătire socio-profesională, nivel socio-cultural etc. Revenind la exemplul de mai sus, eşantionul ar trebui să cuprindă 500 de fete şi 500 de băieţi, câte 250 de elevi pentru fiecare an al ciclului liceal şi, de asemenea, membrii eşantionului trebuie să provină din toate profilele liceelor din ţară.

Teme de reflecţie/ autoevaluare:

1. Care ar fi consecinţele, în cazul în care eşantionul ar cuprinde elevii numai dintr-o anumită zonă geografică a ţării ?
2. Care ar fi consecinţele, în cazul în care eşantionul ar cuprinde elevi numai de la un anumit profil liceal ?

2. Modele de scale folosite în etalonare şi standardizare

Cotele brute (raw score) nu spun nimic în sine, cel mult ele pot fi raportate la media grupului de referinţă. Notele la teză la un anumit obiect de învăţământ la o anumită clasă aparţin practic, tuturor nivelelor, adică vor fi de la 4 (notele sub 4 înseamnă acelaşi lucru, adică tot căzut) la 10. La fel se vor întâmpla lucrurile şi la o altă clasă, de la un liceu dintr-o altă localitate. Vom putea spune că, în interiorul acestei clase, elevul Ionescu a obţinut o notă mai mare decât elevul Popescu (primul a obţinut 9 şi al doilea, 7). În cealaltă clasă, va fi, de asemenea, un elev, să-i spunem Albu, care va obţine nota 9 şi un alt elev, Vasiliu, care va obţine nota 7. Având în vedere notele obţinute, elevul Ionescu este echivalent cu elevul Albu şi elevul Popescu este echivalent cu elevul Vasiliu. Referior la achiziţiile realizate de elevi respectivi nu este obligatoriu ca notele obţinute să le reflecte în mod corect. În cazul în care am avea un test de achiziţie standardizat şi l-am aplica pe elevii noştri nu este exclus ca să apară situaţii care ni s-ar părea paradoxale, respectiv, elevii cu bote mai mici la teze să se plaseze pe poziţii mai bune în ceea ce priveşte rezultatele la testul standardizat decât elevii

Copyright © DEPARTAMENT ID 2008

care au obţinut note mari la teză. Pe de altă parte, deşi notele la teză au fost identice pentru doi elevi, în urma aplicării testului standardizat s-ar putea să nu mai rămână niciun fel de echivalenţă.

Din acest exemplu reiese clar necesitatea construirii unor teste care trebuie să efie etalonate şi standardizate. În procesul etalonării, cotele brute sunt transformate în valori numerice, acestea permiţând raportarea scorului individual al unui subiect, obţinut în urma aplicării testului, la gruparea sau distribuirea rezultatelor obţinute de eşantion.

Etalonarea presupune, în principiu, două operaţii (Radu, I., 1991):

a) alcătuirea sistemuluide norme sau a tabelului de norme pentru grupul de referinţă sau eşantionul de subiecţi (gruparea sau distribuirea în clase a cotelor brute);

b) alcătuirea unui sistem de notare standard în care se convertesc cotele brute ale testului.

În practica psihodiagnostică întâlnim mai ales două categorii de cote transformate, respectiv:

a) cote exprimate, la rândul lor, în decile, centile şi cuartile;

b) cote standardizate, cele care se bazează pe abaterea standard.

Prezentăm, mai întâi, prima categorie de cote transformate.

Sistemul decilajului presupune împărţirea scorurilor sau a datelor brute în 10 grupe sau clase, fiecare dintre ele cuprinzând 10% din efectivul total de subiecţi, exprimat, de regulă, prin litera N. Deci, dacă numărul total al subiecţilor incluşi în eşantion este de 1000, o decilă va cuprinde 10%, respectiv, 100 de subiecţi.

Sistemul centilajului, extrem de asemănător cu sistemul decilajului, presupune împărţirea în 100 de grupe sau clase, fiecare cuprinzând 1% din efectivul total al subiecţilor ce au fost incluşi în eşantion (N). Este de la sine înţeles că acest sistem se poate aplica în cazul în care valoarea lui N este foarte mare.

Sistemul cuartilelor presupune împărţirea în 4 clase sau 4 cuartile, fiecare din ele cuprinzând 25% din N, deci din efectivul total al eşantionului. Cele 4 cvartile sunt următoarele (M = 50% din N):

a) cuartilul superior;

b) cuartilul mediu superior

c) cuartilul mediu inferior

d) cuartilul inferior

Folosirea tuturor celor 3 sisteme presupune parcurgerea unor etape asemănătoare, respectiv:

1. aplicarea testului pe întreg eşantionul (N);
2. ordonarea scorurilor sau rezultatelor, de la valoarea cea mai mare sau cea mai mică a performanţelor;
3. stabilirea frecvenţelor, adică a numărului de subiecţi care obţin aceeaşi performanţă (fa);
4. stabilirea frecvenţelor cumulate sau însumate (fc), încât Σ fc = N.

Vom încerca să exemplificăm folosind rezultatele obţinute la testul *Raven Standard,* aplicat pe 448 de candidaţi de la un examen de admitere (Clinciu, A.I., 2005). Rezultatele obţinute sunt incluse, mai întâi, în tabelul de mai jos:

x	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**
f_a	0	0	1	0	1	3	0	0	1	0	1	0	0	1	0	3	1	2
f_c	0	0	1	1	2	5	5	5	6	6	7	7	7	8	8	11	12	14

Copyright © DEPARTAMENT ID 2008

x	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**
f_a	3	4	7	3	2	11	12	15	18	24
f_c	17	21	28	31	33	44	56	71	89	113

x	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**	**55**
f_a	20	28	28	36	37	32	29	35	23	28
f_c	133	161	189	225	262	294	323	358	381	404

x	**56**	**57**	**58**	**59**	**60**
f_a	22	13	4	5	0
f_c	426	439	443	448	448

Vom proceda, în continuare, folosind rezultatele incluse în acest tabel, la alcătuirea etaloanelor în cuartile şi decile (etalonul în centile, aşa după cum am mai precizat, este foarte asemănător cu etalonul în decile.

Etalonul în cuartile. Fiecare cuartil cuprinde 25% din N. Deci primul cuartil (cel inferior) are înaintea sa 25% din subiecţi, adică 112 (448 x 25/100). Căutăm pe şirul frecvenţelor cumulate (f_c) cea mai apropiată valoare de 112. Aceasta corespunde cotei brute **x** = 45. În consecinţă, acest cuartil va cuprinde valorile lui **x** de la 0 la 45. Cuartilul 2 (mediu inferior) împarte pe din două efectivul, deci are înaintea sa 50% din efectivul subiecţilor (N), adică 224 (448 x 50/100). Valoarea cea mai apropiată de aceasta este 225 şi ea corespunde cotei brute = 49. Drept urmare, cuartilul 2 va cuprinde valorile lui **x** între 46 şi 49. Cuartilul 3 (mediu superior) are înaintea sa 75% din N, adică 336 (448 x 75/100). Valoarea cea mai apropiată este 353 şi ea corespunde cotei brute = 53. Deci cuartilul 3 va cuprinde valorile lui **x** între 50 şi 53 . În sfârşit, prin diferenţă, cuartiul 4 **x** între 54 şi 60.

Finalmente, etalonul în patru clase (sau cuartile) va arăta în felul următor:

Cuartilul	Clase	Procente	
		Teoretice	Cumulate
Cuartilul 1	0 - 45	25	25
Cuartilul 2	46 - 49	25	50
Cuartilul 3	50 - 53	25	75
Cuartilul 4	54 - 60	25	100

Etalonul în decile. Vor fi folosite tot datele rezultate în urma aplicării testului Raven. Primul decil are 10 % din N înaintea sa, deci, 44,8 (448 x 10/100). În urma rotunjirii, se obţine 45. Căutând pe şirul frecvenţelor cumulate vom găsi valoarea cea mai apropiată, care este 44 şi care corespunde cotei brute, respectiv, valorii lui x = 41. Asta înseamnă că primul decil va cuprinde valori ale lui x între 0 şi 41. Al doilea decil are 20% din efectivul total al subiecţilor (N) înaintea sa, ceea ce înseamnă 89,6 (448 x 20/100) şi rotunjit, 90. Urmărind pe şirul frecvenţelor cumulate, vom vedea că valoarea cea mai apropiată este 89 şi ei îi corespunde o valoare a lui x = 44. deci, al doilea decil va cuprinde valori ale lui x între 42 şi 44. Aşa se procedează şi pentru celelalte decile iar, în final, etalonul în decile va arăta în felul următor:

Copyright © DEPARTAMENT ID 2008

Decilul	Clase	Procente	
		Teoretice	Cumulate
1	0 - 41	10	10
2	42 - 44	10	20
3	45 - 46	10	30
4	47 - 48	10	40
5	49	10	50
6	50	10	60
7	51 - 52	10	70
8	53	10	80
9	54 - 55	10	90
10	56 - 60	10	100

Al doilea sistem de cote transformate presupune cotele standardizate (Radu, I., 1991), care înseamnă tot împărţirea în clase dar clase normalizate. Este vorba tot de un calcul procentual, însă efectivul total al eşantionului nu se împarte în clase sucesive egale (10%, 20% etc.), ci în clase care cuprind procente conturate plecând de la frecvenţele distribuţiei normale. Regula este ca numărul claselor să fie un număr impar: 5, 7, sau 9 clase, procentele fiind, desigur, diferite de la un sistem la altul şi anume:

a) pentru 5 clase: 6,7%; 24,2%; 38,2%; 24,2%; 6,7%;

b) pentru 7 clase: 4,8%; 11,1%; 21,2%; 25,8%; 21,2%; 11,1%; 4,8%;

c) pentru 9 clase: 4,0%; 6,6%; 12,1%; 17,5%; 19,6%; 17,5%; 12,1%; 6,6%; 4,0%.

În practică sistemul stanine este cel mai des folosit, de aceea îl vom prezenta şi noi mai în detaliu. Vom face trimitere tot la rezultatele obţinute în urma aplicării testului Raven (Clinciu, A., 2005).

Stanina 1 are sub ea 4% din efectiv, deci 17,92 (448 x 4/100), deci, rotunjit, 18. Căutând pe şirul frecvenţelor cumulate vom constata că valoarea cea mai apropiată este 17 şi acesteia îi corespunde cota brută (valoarea lui **x**) = 36. Stanina 2 are înaintea ei 10,6 procente din efectiv (4,0% + 6,6%), adică 47,468 (448 x 10,6), ceea ce înseamnă, rotunjit, 47. Căutând pe şirul frecvețelor cumulate, constatăm că cea mai apropiată valoare este 44, căreia îi corespunde cota brută = 41. Stanina 3 are înaintea sa 22,7 procente din efectiv (10,6 + 12,1), adică 101,696 (448 x 22,7/100) şi, deci, rotunjit, 102. Căutând pe şirul frecvețelor cumulate, constatăm că cea mai apropiată valoare este 113, căreia îi corespunde cota brută = 45. Ş.a.m.d. În final, etalonul va arăta în felul următor:

Stanina	Clase	Procente	
		Teoretice	Cumulate
1	0 - 36	4,0	4,0
2	37 - 41	6,6	10,6
3	42 - 45	12,1	22,7
4	46 - 48	17,5	40,2
5	49 - 50	19,6	59,8
6	51 - 53	17,5	77,3
7	54 - 55	12,1	89,4
8	56	6,6	96,0
9	57 - 70	4,0	100,0

Copyright © DEPARTAMENT ID 2008

Teme de reflecție/autoevaluare:

1. Arătați cum va arăta etalonul, în cazul în care folosiți o scală cu 5 trepte;
2. Arătați cum va arăta etalonul în cazul în care folosiți o scală cu 7 trepte. Pentru ambele întrebări veți face apel tot la rezultatele obținute în urma aplicării testului Raven.

Întrebări de autoevaluare:

1. Ce înseamnă revizia testelor ?
2. Prin câte revizii a trecut Bateria de inteligență Stanford-Binet ?
3. Prin câte revizii a trecut Bateria de inteligență WAIS ?
4. Când se impune revizia unui test ?
5. Care sunt etapele elaborării şi construirii unui test nou ?
6. Care sunt principalele cerințe în elaborarea unui test nou ?
7. Ce înseamnă reprezentativitatea eşantionului ?
8. Câte categorii de cote transformate cunoaşteți ?
9. Care este diferența dintre sistemul decilelor şi sistemul centilelor ?
10. Care este specificul sistemului cuartilelor ?
11. Care sunt principalele etape ce trebuie parcurse când folosim oricare dintre cele 3 sisteme ?
12. Ce înseamnă cote standardizate ?
13. Care este specificul sistemului stanine ?
14. De ce etaloanele bazate pe cote standardizate au un număr impar de clase ?
15. Care este relația dintre frecvența absolută şi frecvența cumulată ?

BIBLIOGRAFIE:

1. ALBU, M. (1998). *Construirea şi utilizarea testelor psihologice*. Cluj: Editura Clusium.
2. ALBU, M., PITARIU, H. (1993). *Proiectarea testelor de cunoştințe şi examenul asistat de calculator*. Cluj-Napoca: Editura Casa cărții de ştiință.
3. COHEN, J.R., SWERDLIK, E.M. (2005). *Psychological Testing and Assessment. An Introduction to Tests and Measurement*. McGraw-Hill International Edition.
4. CLINCIU, A.I. (2005). *Psihodiagnostic*. Braşov: Ed. Universității „Transilvania"
5. FRIEDENBERG, L. (1995). *Psychological Testing. Design, Analysis, and use.* Allyn & Bacon.
6. KAPLAN, M.R., SACCUZZO, P.D. (2005). *Psychological Testing. Principles, Applications, and Issues*. Thomson Wadsworth.
7. MINULESCU, M. (2003). *Teorie şi practică în psihodiagnoză*. *Testarea intelectului*. Bucureşti: Editura Fundației România de Mâine.
8. RADU, I. (coord.) (1991). *Metodologie psihologică de analiză a datelor*. Cluj-Napoca: Editura Sincron.
10. STAN, A. (2002). *Testul psihologic.Evoluție, consatrucție, aplicații*. Iaşi: Polirom.

Copyright © DEPARTAMENT ID 2008